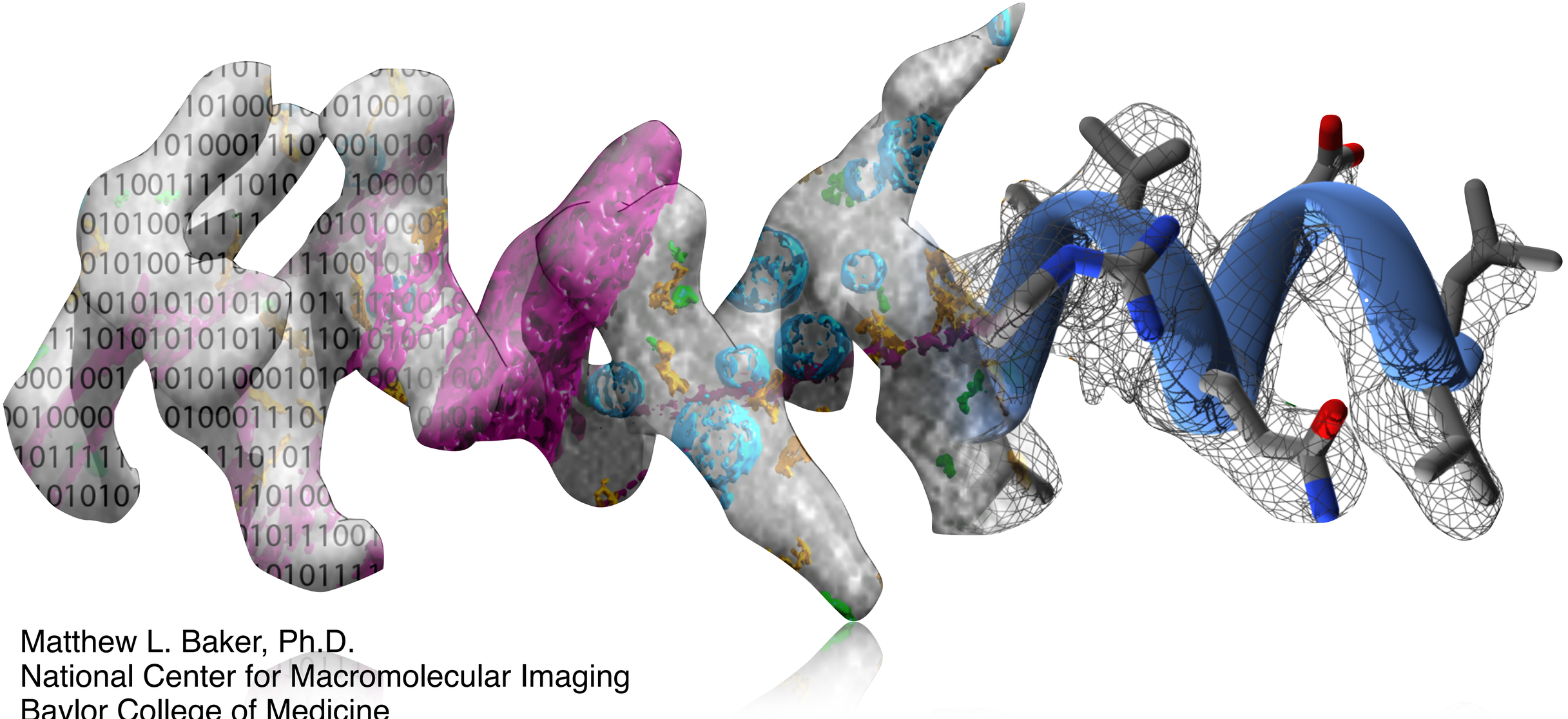
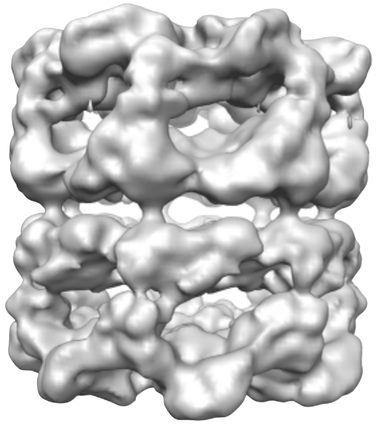
Analysis and Modeling of a Cryo-EM Density Map



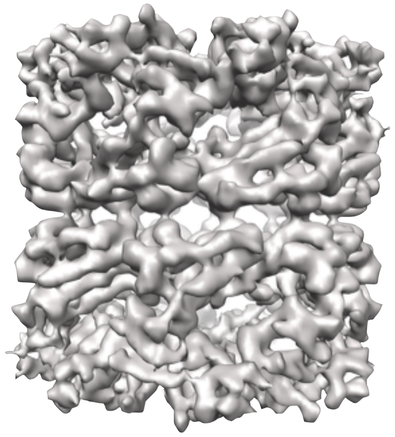
Matthew L. Baker, Ph.D.
National Center for Macromolecular Imaging
Baylor College of Medicine

Resolution in Cryo-EM: GroEL

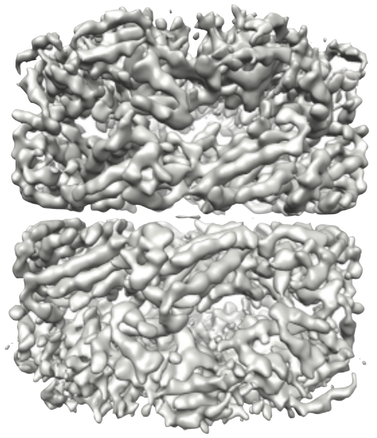
EMDB 5143: 18Å



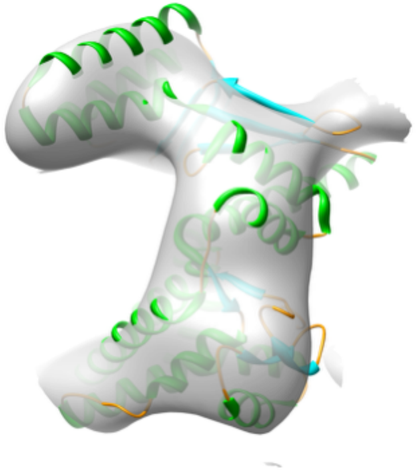
EMDB 1042: 10Å

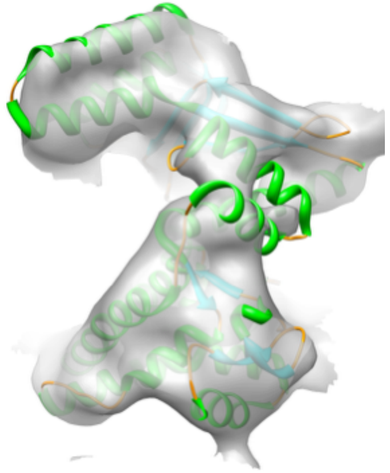


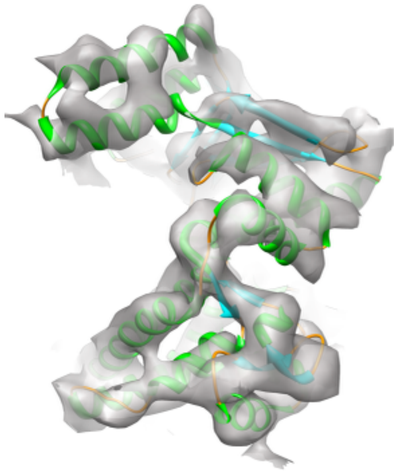
EMDB 1200: 8Å

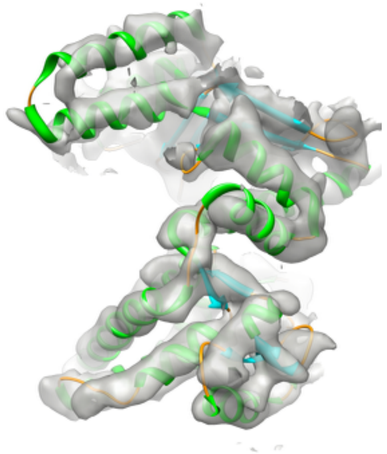


EMDB 5001: 4Å

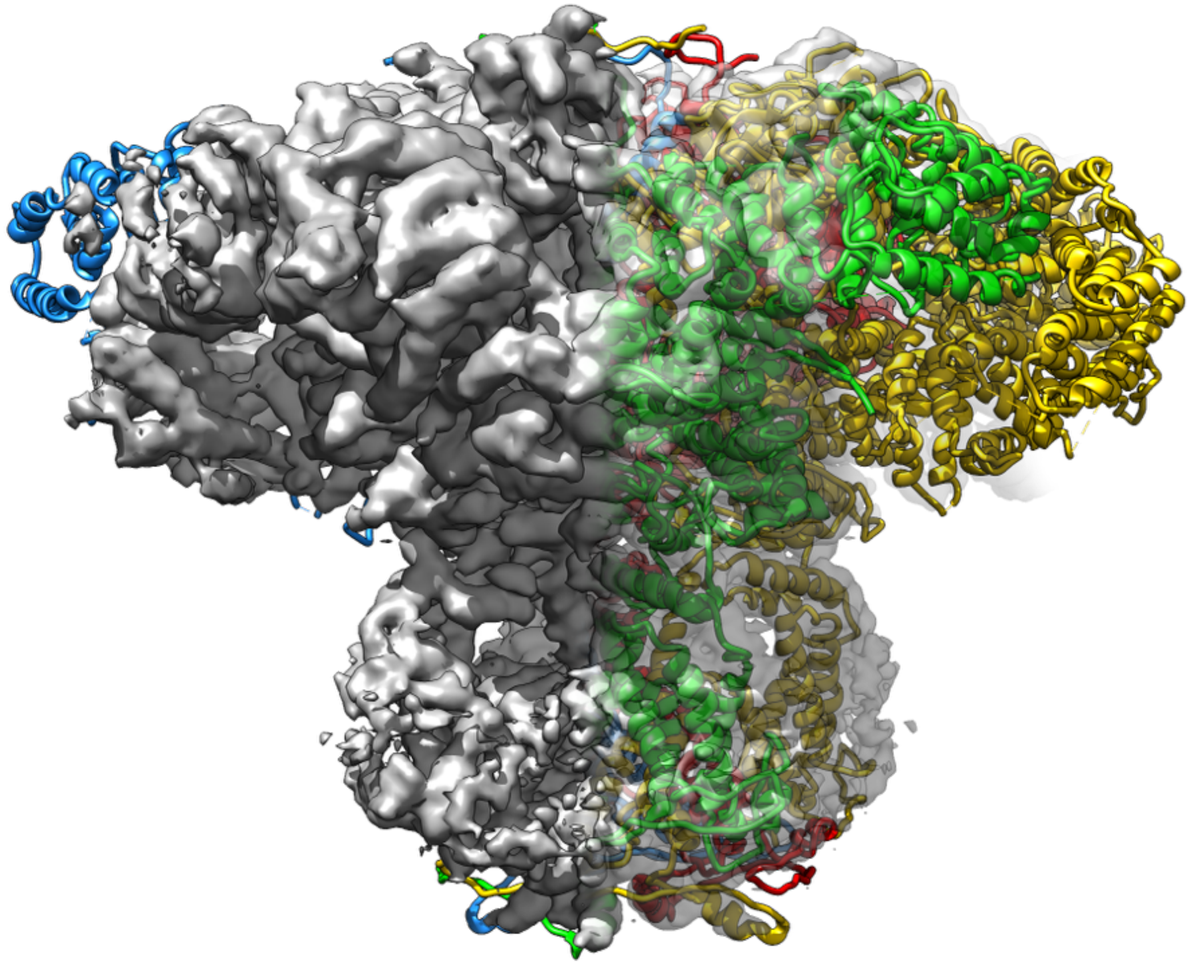


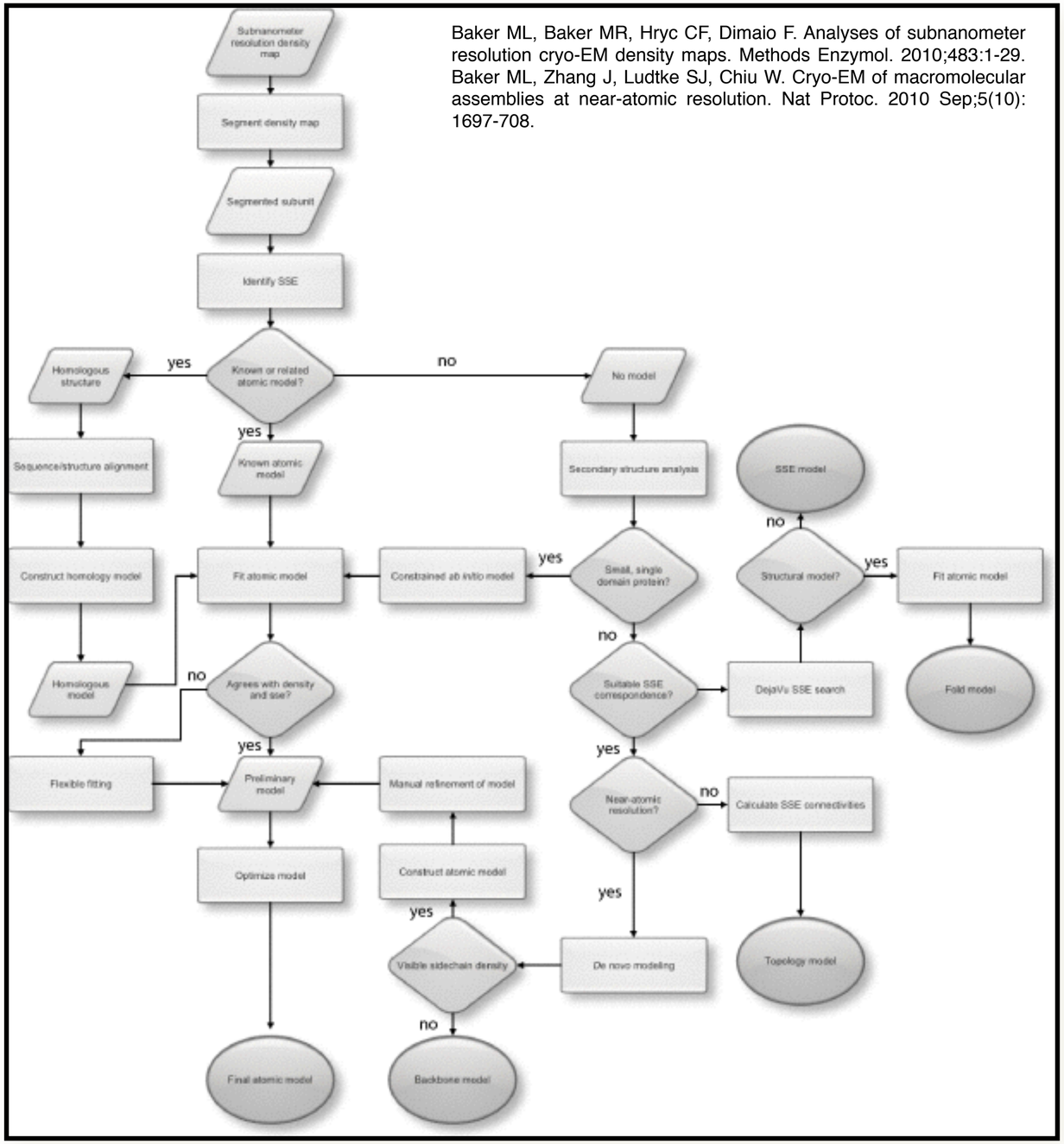






Modeling Pathway





Baker ML, Baker MR, Hryc CF, Dimaio F. Analyses of subnanometer resolution cryo-EM density maps. Methods Enzymol. 2010;483:1-29.
Baker ML, Zhang J, Ludtke SJ, Chiu W. Cryo-EM of macromolecular assemblies at near-atomic resolution. Nat Protoc. 2010 Sep;5(10): 1697-708.

Tools

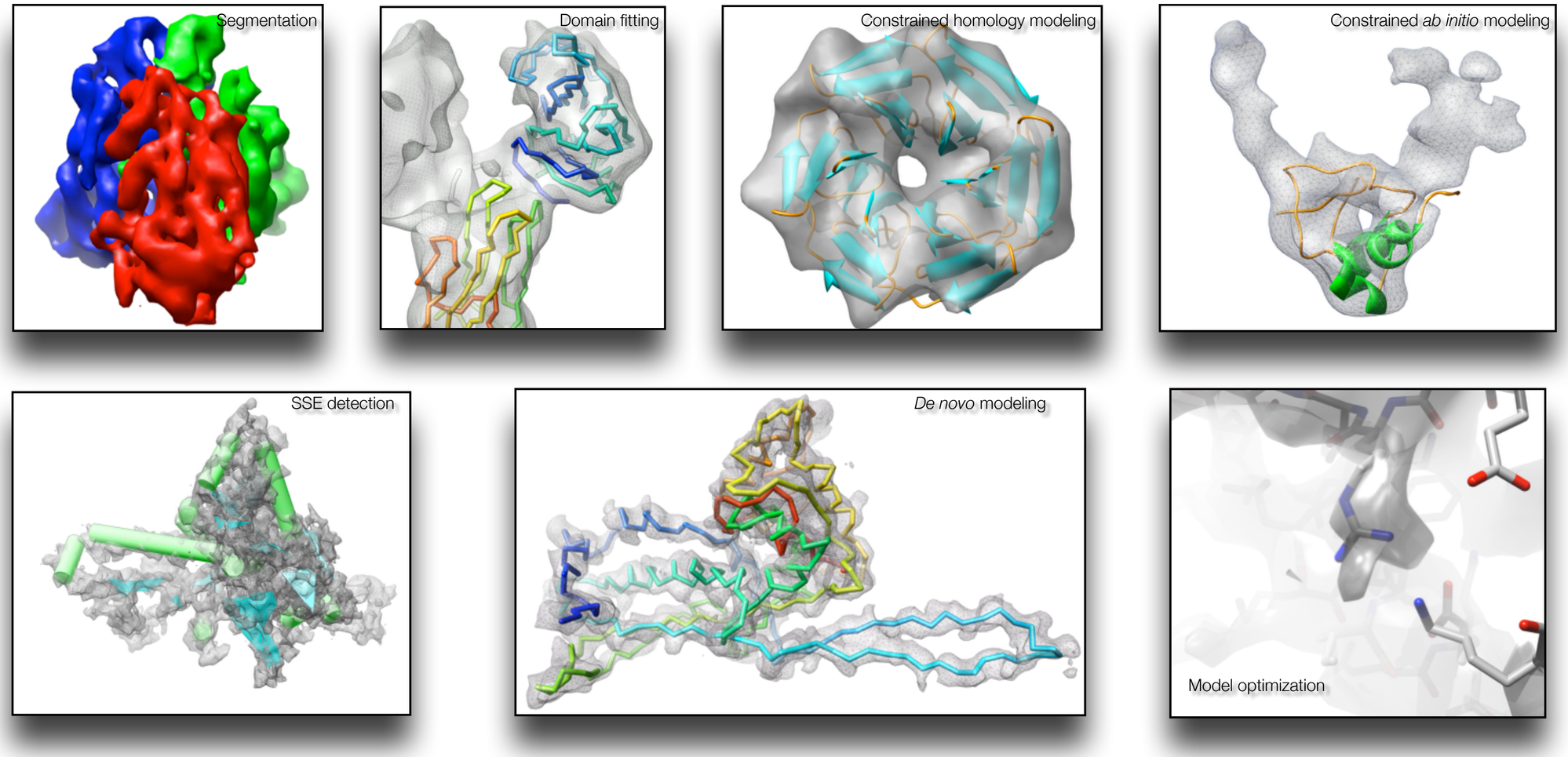
Density Manipulations

- *Filtering*: Mask or de-noise the density map
- *Segmentation*: Identifying and isolating single subunits or domains in the density
- *Feature Recognition*: Identify and characterize density features

Modeling Tools

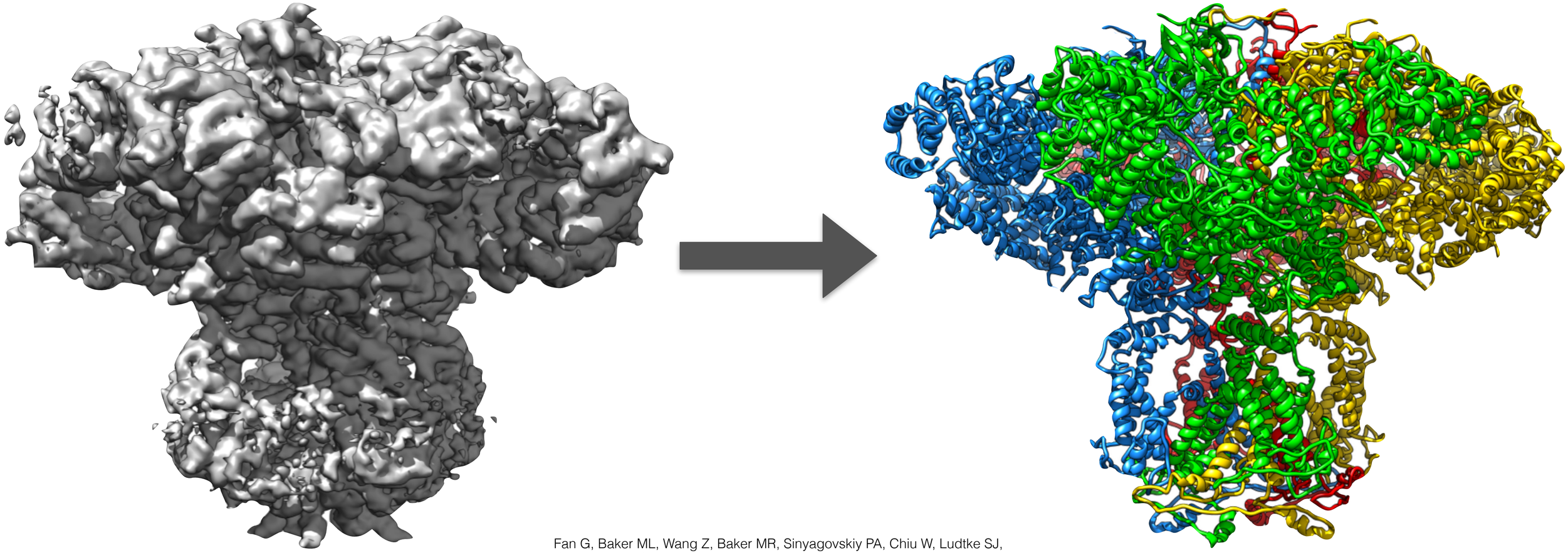
- *Fitting*: Localizing a known structure within the density map
- *Constrained Modeling*: Generating computational models in the context of density
- *Feature Recognition*: Identify density features
- *De novo modeling*: Model building without a template
- *Optimization*: Refinement of a model in the density map

Analyzing Cryo-EM Density Maps



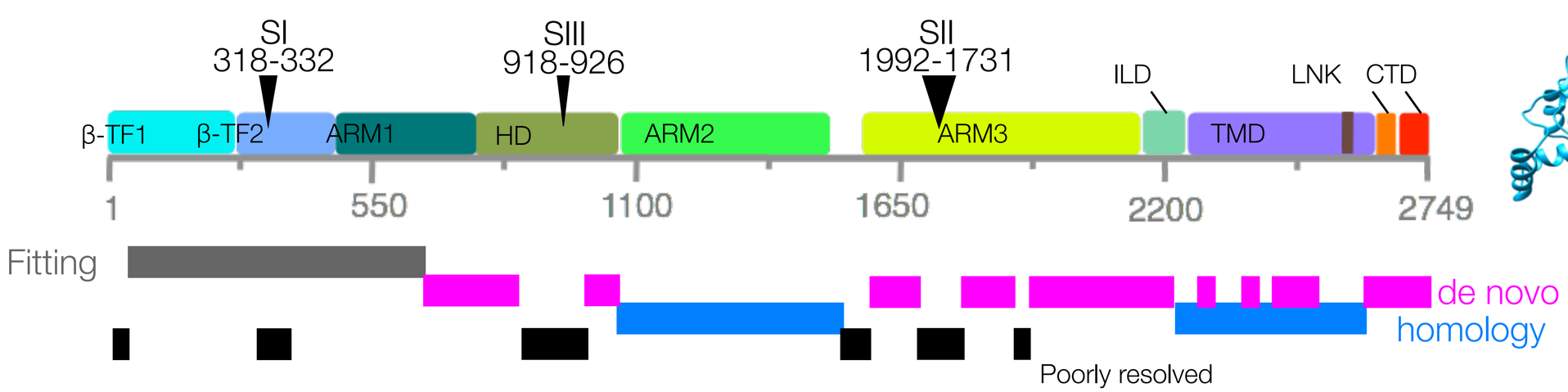
Baker, M. L., Baker, M. R., Hryc, C. F., and Dimaio, F. (2010). Analyses of subnanometer resolution cryo-EM density maps. Methods Enzymol 483, 1-29.

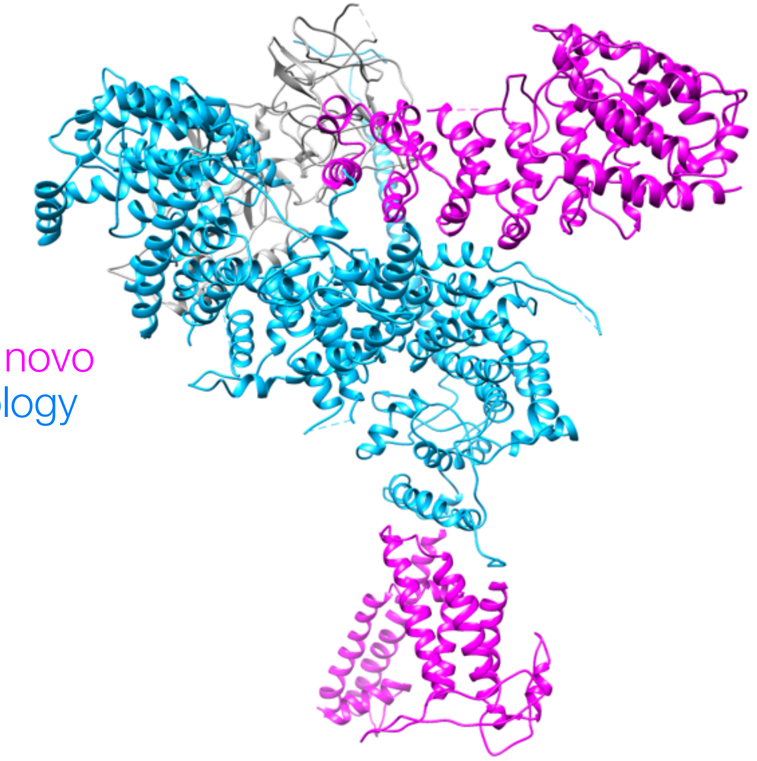
From Cryo-EM Density Maps to Models



Fan G, Baker ML, Wang Z, Baker MR, Sinyagovskiy PA, Chiu W, Ludtke SJ, Serysheva II. Gating machinery of InsP(3)R channels revealed by electron cryomicroscopy. Nature. 2015 Oct 12.

Building an IP3R1 Subunit





Fitting Known Structures

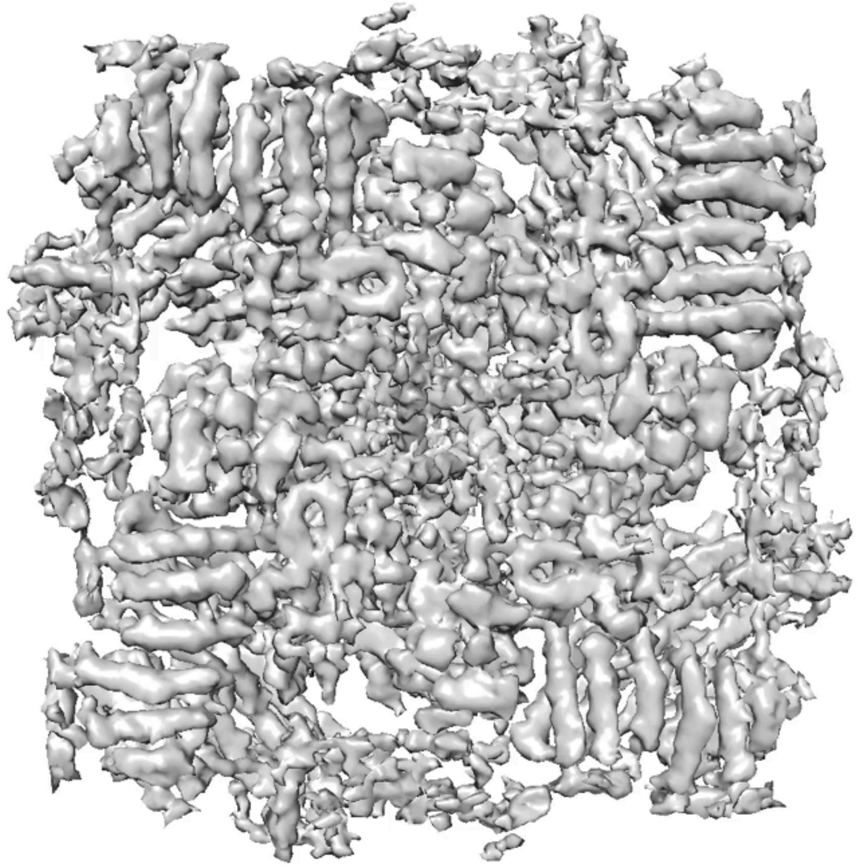
- Known structures/models (probes) for one or more of the components can be fit to a cryoEM density map (target)
  - Interactive and automated programs
  - Assess fit to map (Methods include correlation, atom inclusion/exclusion and clashes avoidance
- Sequential or simultaneous fit of multiple models
- Rigid body vs flexible fit
- Map resolvability key in determining quality of fit

Fitting Software

- Rigid body
  - ➡ ***Foldhunter***, Gorgon, EMFIT, ***UCSF Chimera***, CoAn, ***Situs***, UROX
- Flexible
  - ➡ Gorgon, MDFF, EMFF, ***Phenix***, ***FlexEM***, MDFit, MVP-Fit, Direx, Norma

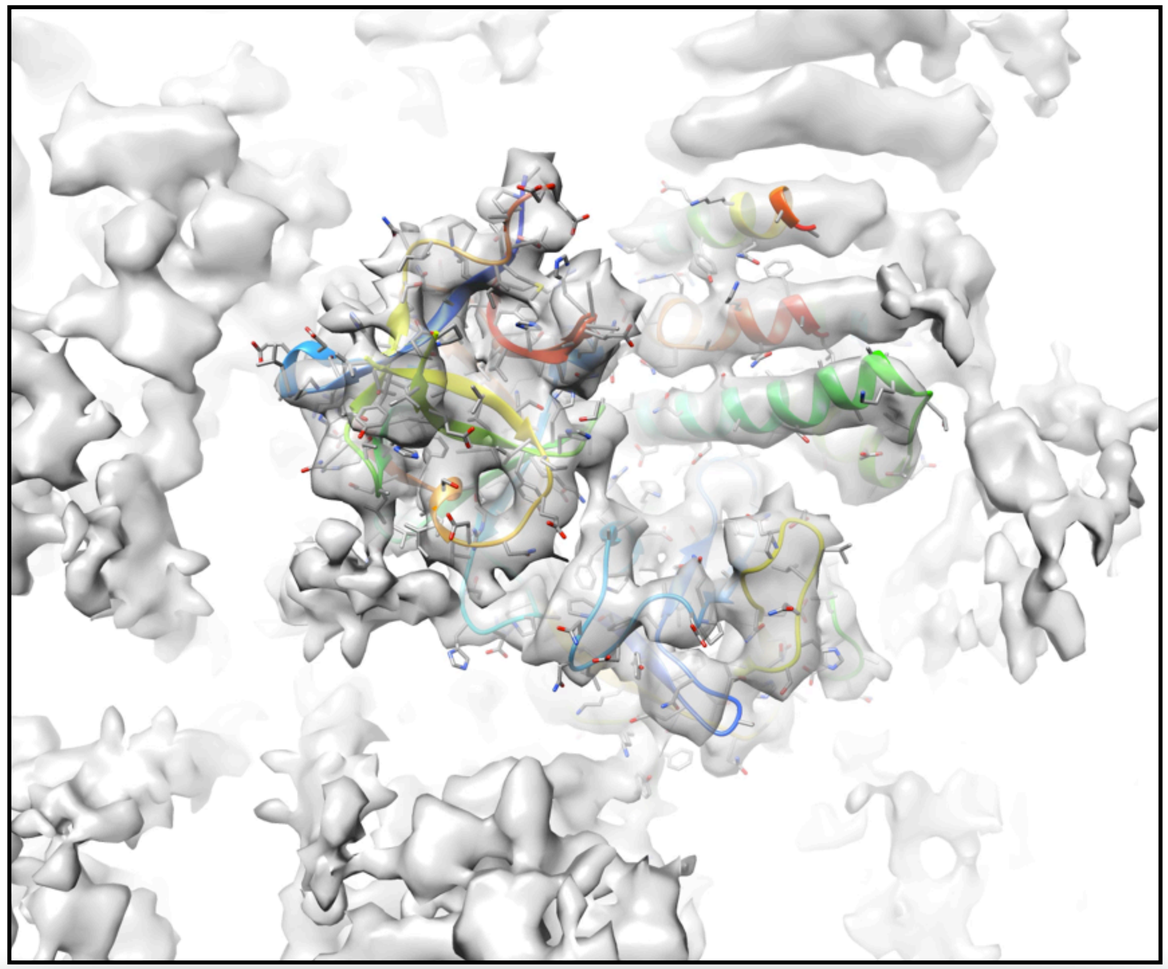
IP3R1: Fitting a Known Structure

- X-ray structure of the N-terminal domain
  - 4 overlapping structures for residues 7-587
  - Includes both ligand-bound and un-bound structures
  - Corresponding N-terminal domain of related RyR1 also available and can “fill-in” missing portions of IP3R1 model
- Fit to entire density map with Chimera, Foldhunter and Situs
  - ~2Å RMSD between different models and fits



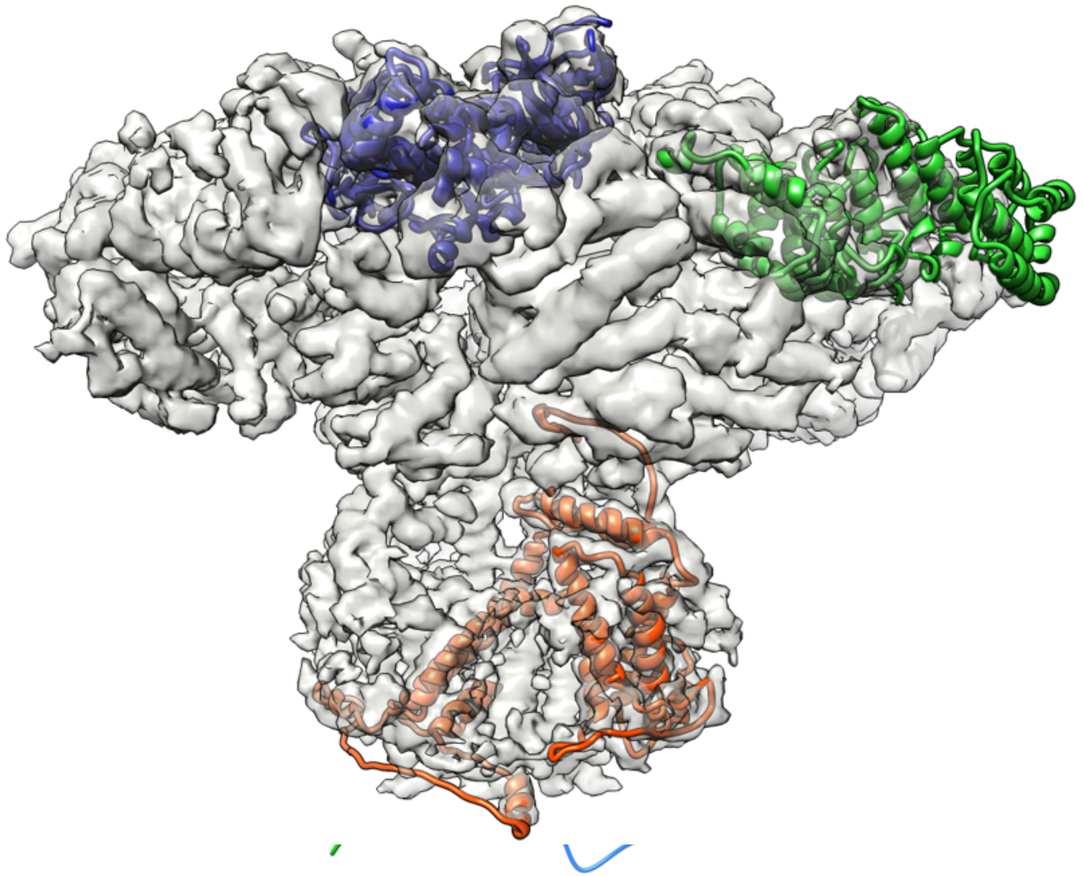
IP3R1: Flexible Fitting

- ➡ Models refined to density map using real-space refinement tools in Phenix and FlexEM
  - ➡ 2.4Å RSMD from original rigid-body fit structure



Density Constrained Modeling

- Sequence-based searches reveal two regions of sequence homology using Phyre and RaptorX
  - Armadillo repeat domain (ARM2, ~1050-1500)
  - Transmembrane domain (TMD, ~2200-2600)
- Homology models fit with Chimera, Foldhunter and Situs
- Models refined against density map with Rosetta and Phenix



Feature Recognition

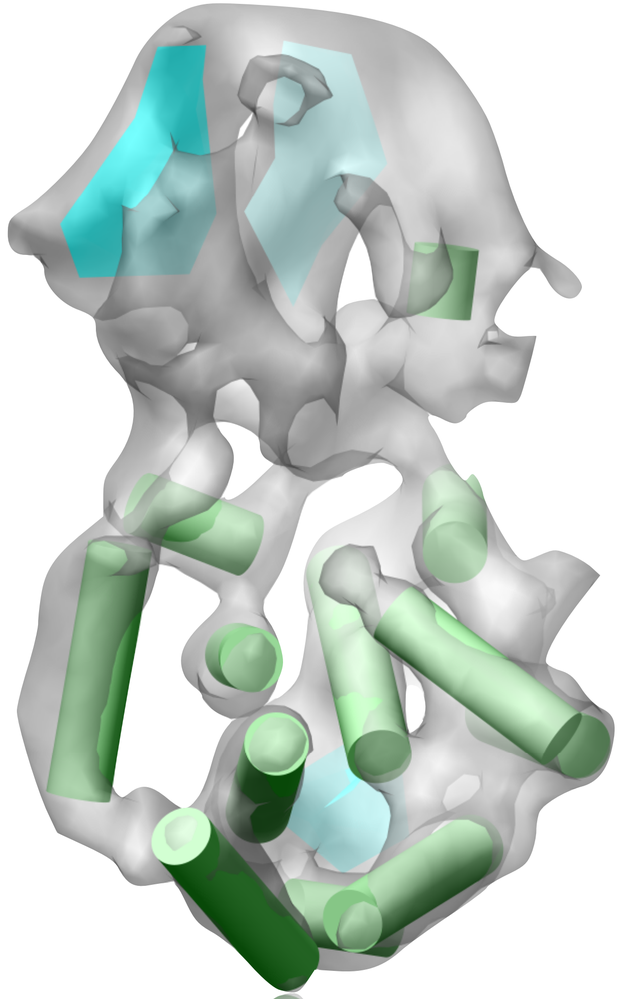
- Localization of individual secondary structure elements within a density map
  - Interactive and automated programs
- Can provide a simple topological model

Secondary Structure Detection Software

- Helixhunter, SSEHunter, StrandTwister, HelixTracer, sheetminer, sheettracer, ***Gorgon***, Pathwalker

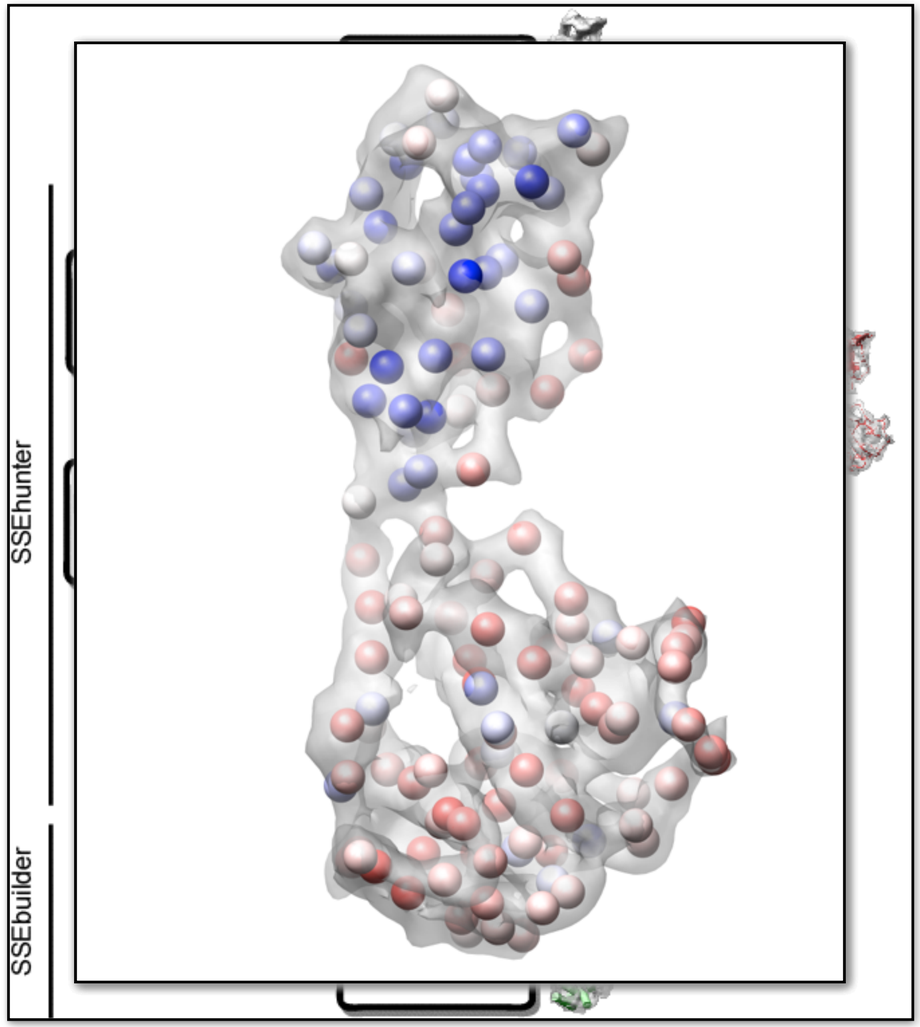
Detecting Secondary Structure Elements

- SSEHunter: guided identification of alpha helices and beta sheets at intermediate resolutions
- Scoring based on correlation, skeletonization and local geometry
- >95% helix (2+ turns) detection accuracy
- >99% detection of 3+ stranded sheets



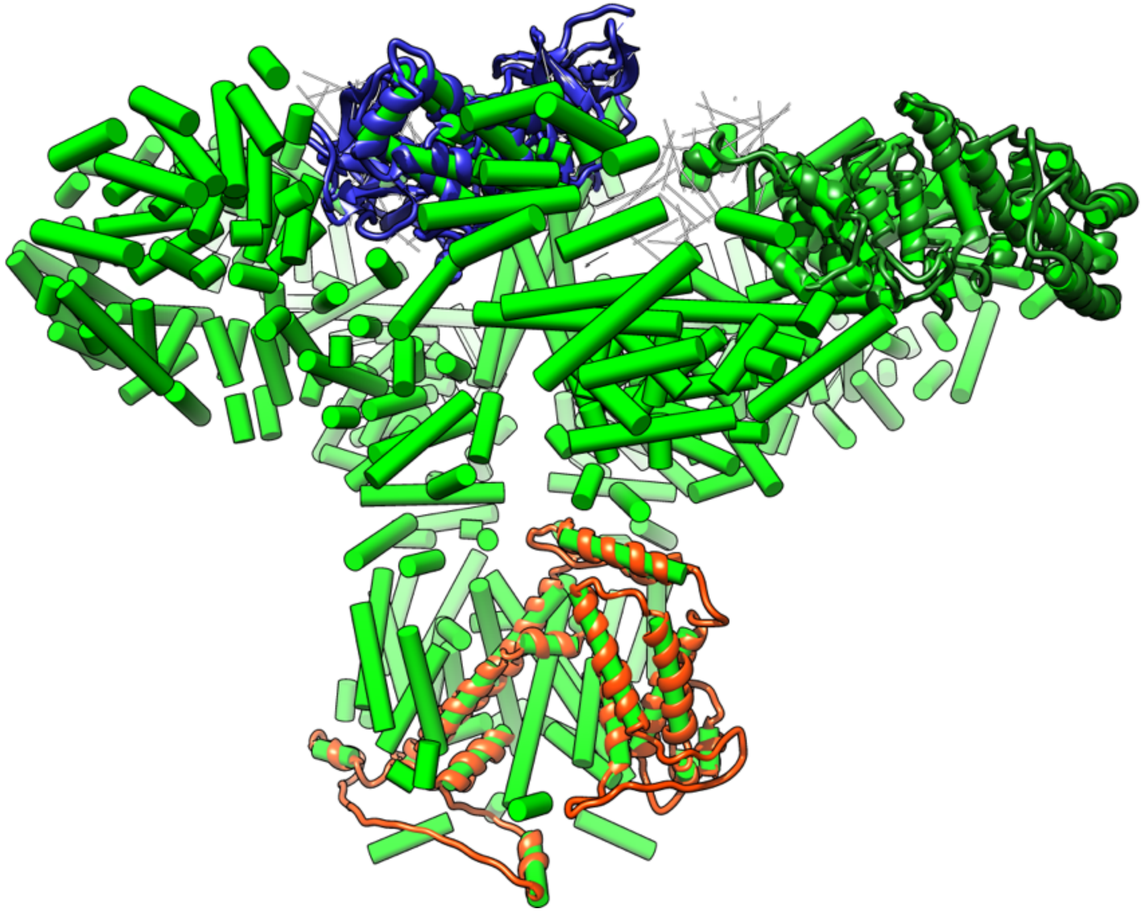
Baker, M.L., Ju, T., Chiu, W. (2007) Identification of Secondary Structure Elements in Intermediate Resolution Density Maps. Structure (15), p 7-19.

SSEHunter Methodology

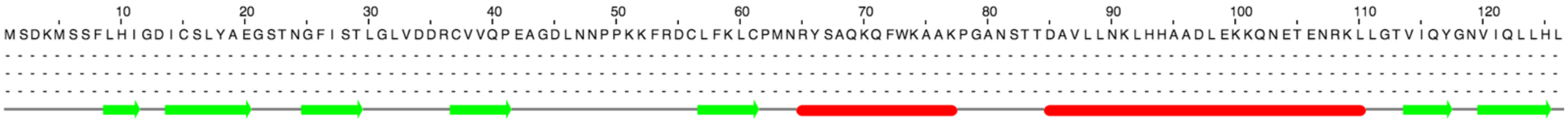


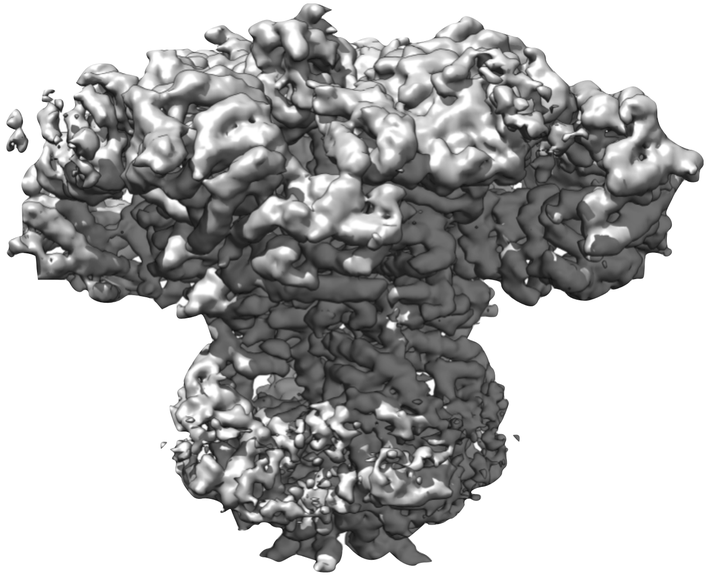
IP3R1: SSE Detection

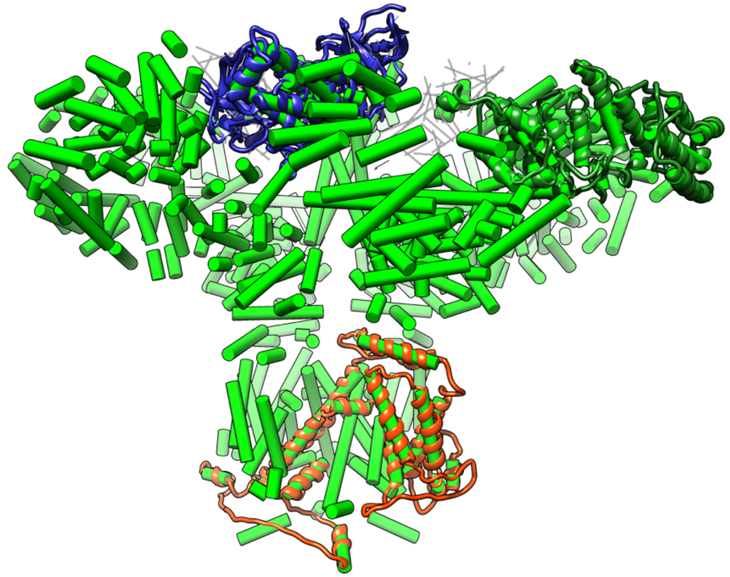
- 82 “good” helices identified in the density of map with SSEHunter
- Two sheet domains
- SSEHunter helices matched helices in fitted structures
- >90 helices per IP3R1 monomer predicted in sequence



From Map to Model

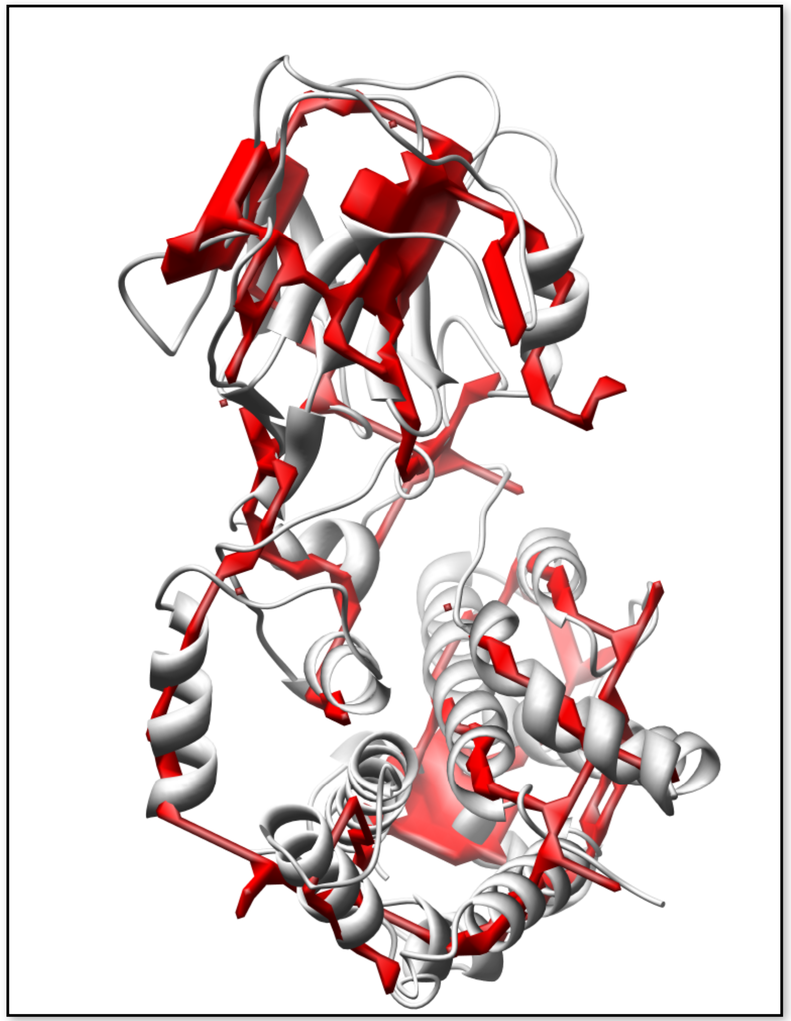






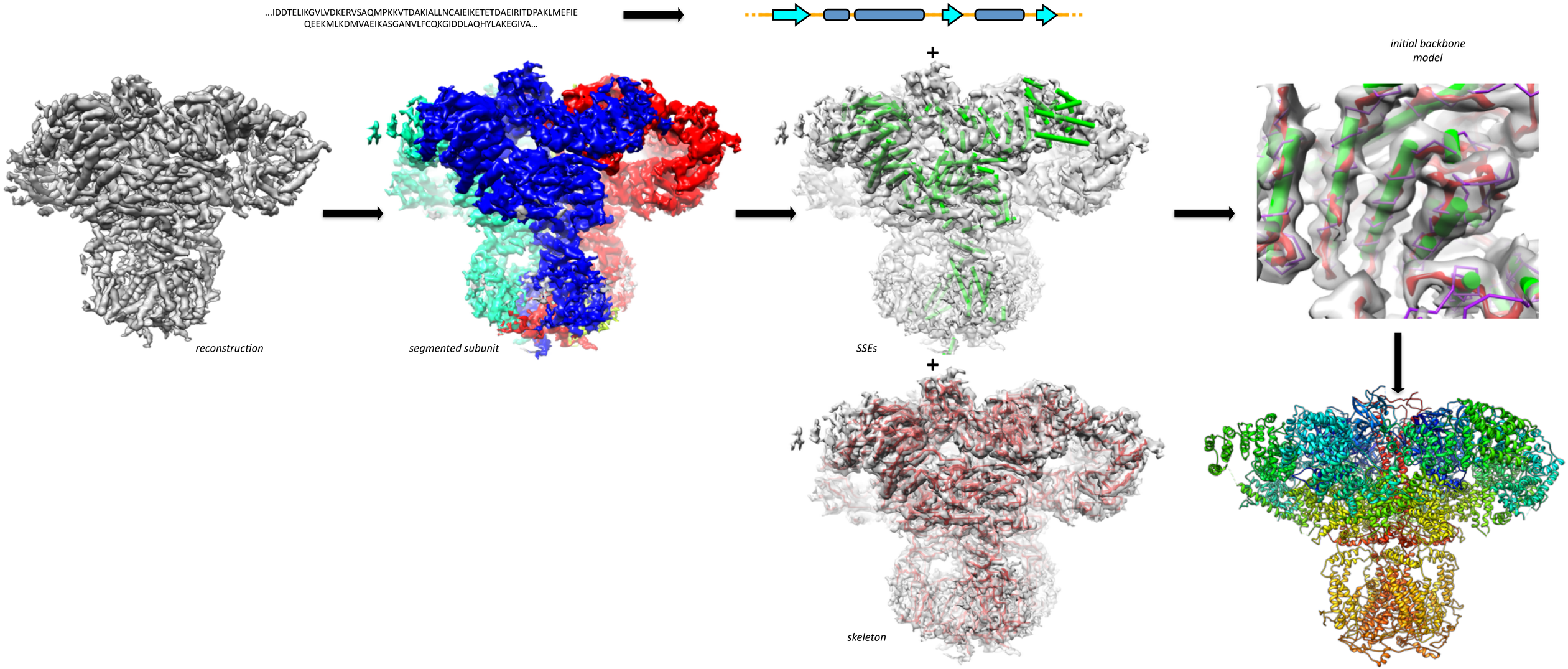
Building SSE Connections

- Density skeletonization - compact geometric representation of a volume
- Feature preserving
  - Sheets are represented as flat surfaces
  - Helices and loops are represented as curves
- Topology preserving
  - Maintains density connectivity
  - minimizes branches and breaks



Ju, T., Baker, M.L., Chiu, W. (2007). Computing a Family of Skeletons of Volumetric Models for Shape Description. Computer-AIDED Design (39), p 352-360.

Step 3: *de novo* Modeling



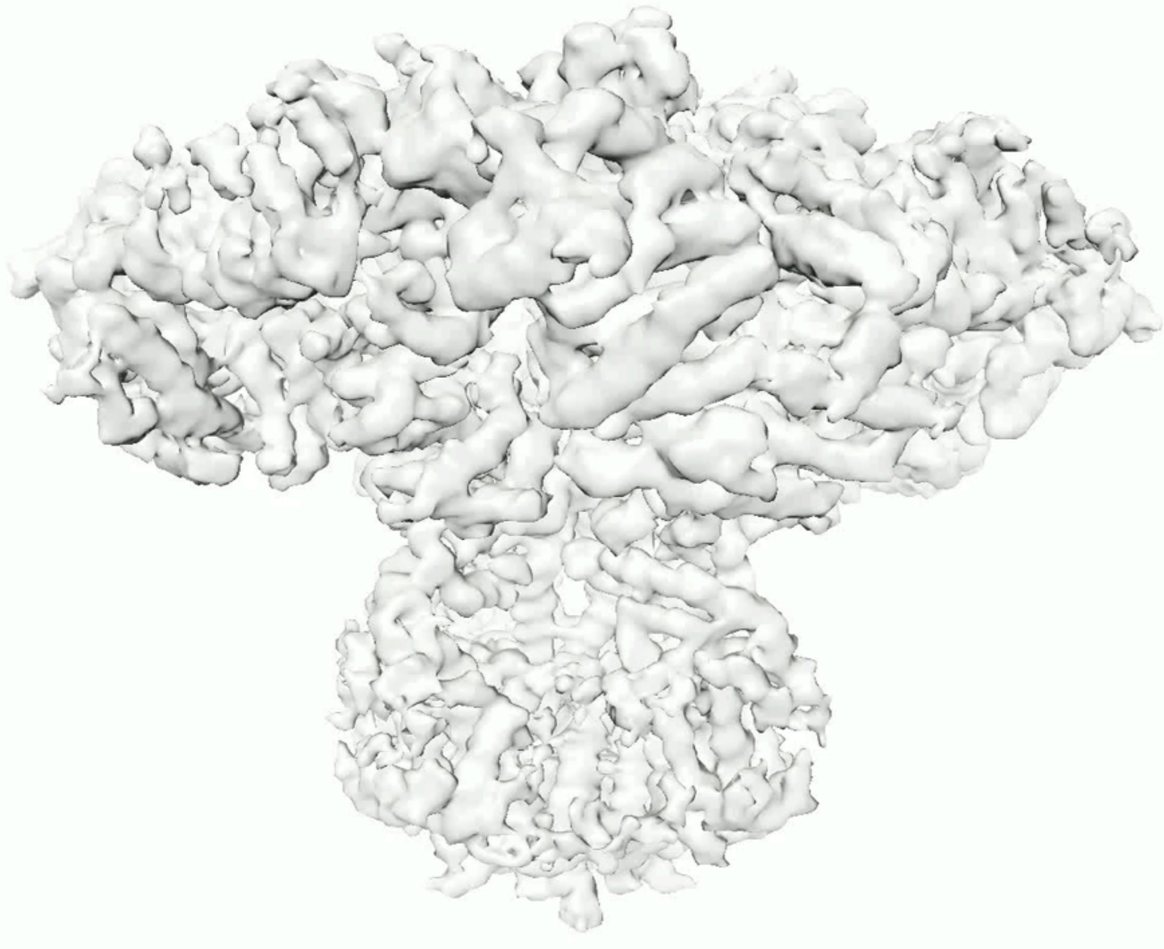
IP3R1 *de novo* Model

➡Fitted models and SSE correspondence serve as anchor points

➡Complete topological model for ~85% of the IP3R1 monomer

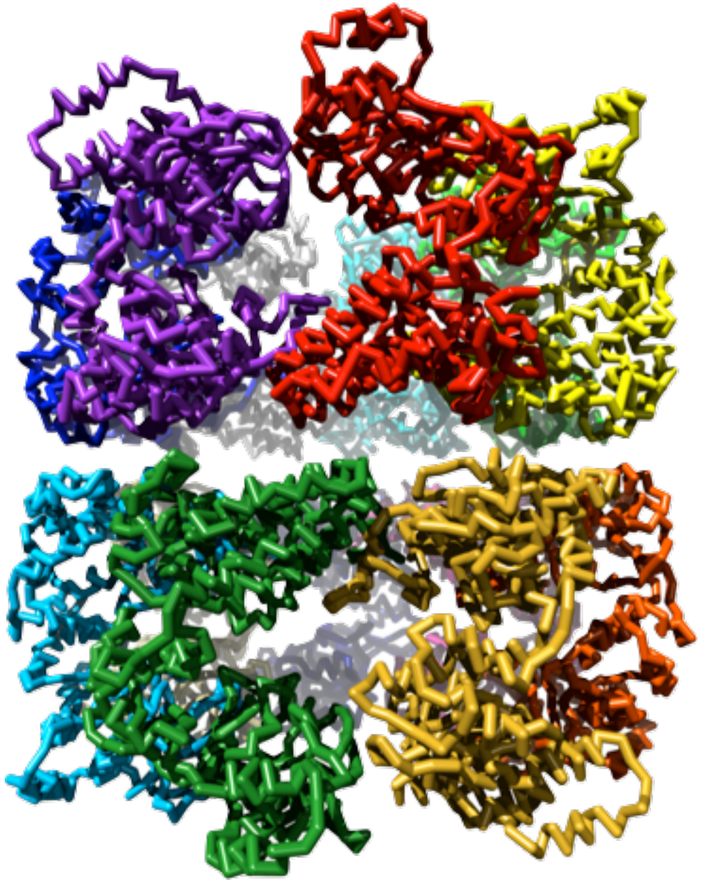
- ➡ Density for the loops at the three splice variants are missing
- ➡ Small loops missing in N-terminal domain
- ➡ Missing connection between ARM2 and ARM3 ~(100aa)

➡ Model consists of both full-atom domains (Fitted and homology models) and C-alpha only domains (de novo)



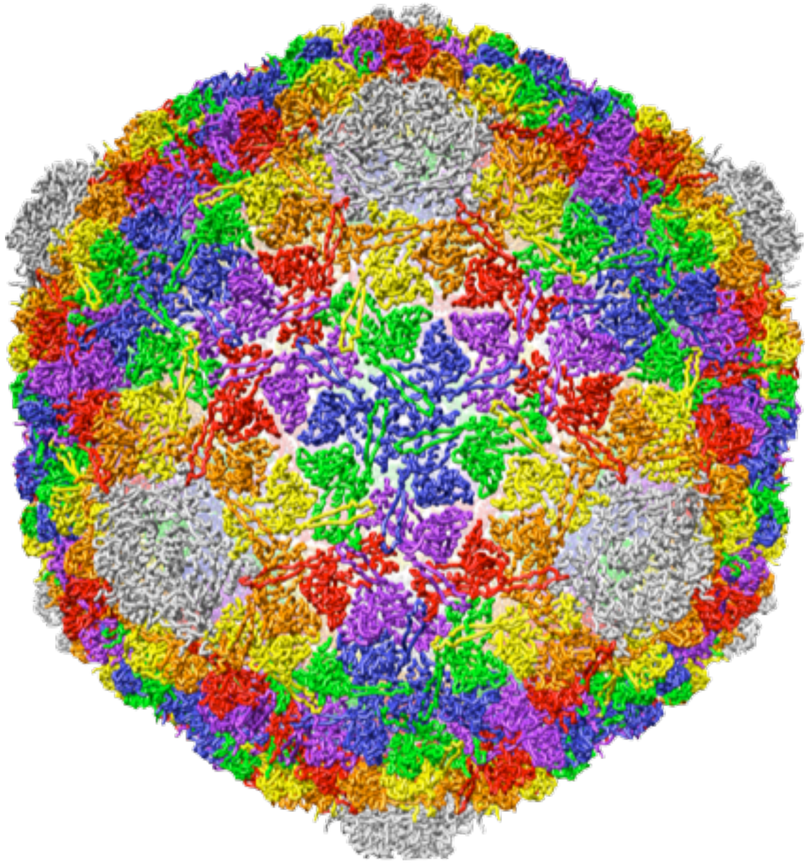
Model Building

4.2Å resolution GroEL



Ludtke, S.J.*, Baker, M.L.*, Chen, D.H. Song, J.L., Chuang, D.T., Chiu, W. (2008) Structure (16), p 441-448.

4.5Å resolution ε15

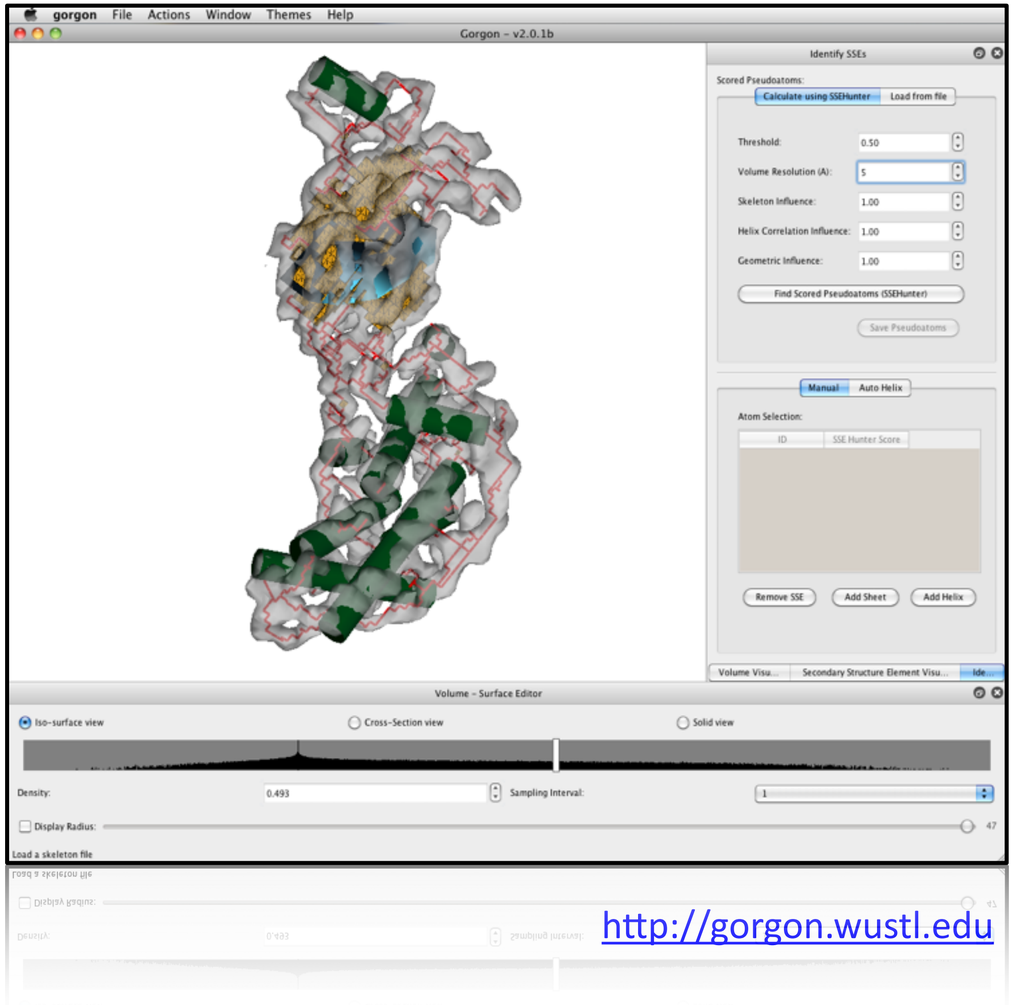


Jiang, W.*, Baker, M.L.*, Jakana, J Weigele, P.R., King, J., Chiu W. (2008) Nature (451), p 1130-1135.

Gorgon

Interactive molecular modeling toolkit for intermediate resolution density maps focused on providing a simple and efficient framework for de novo modeling

- ➡Released Version 1 in Dec. 2008, Version 2.2b released in 2015
- ➡Annual workshops and trainings
- ➡On-line videos and tutorials with sample data
- ➡Cross platform (Windows, Linux, OS X 10.5+)

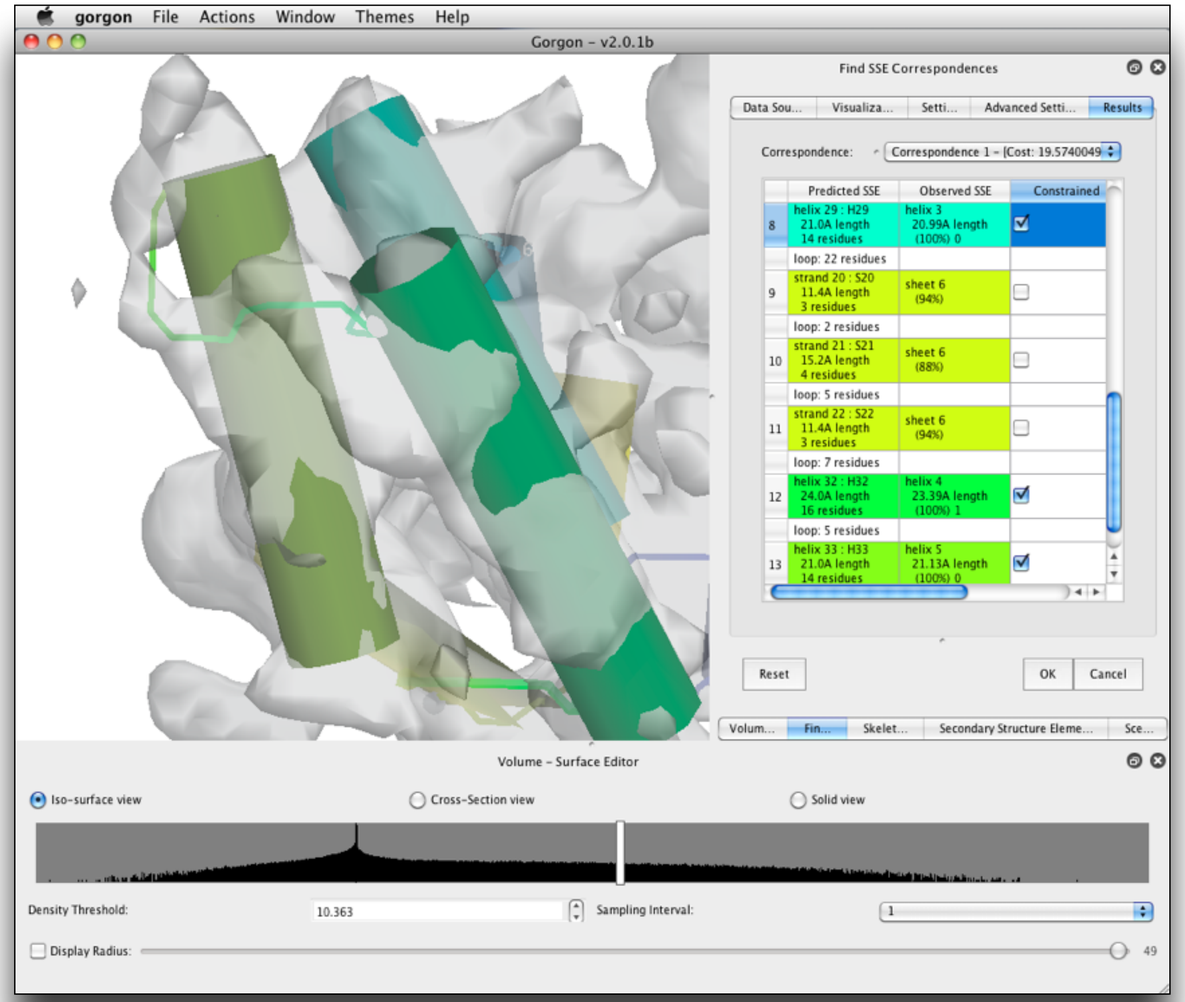


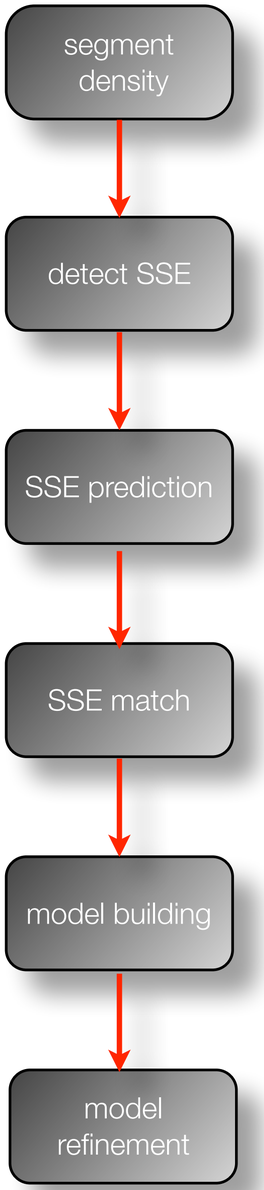
http://gorgon.wustl.edu

Baker ML, Abeysinghe SS, Schuh S, Coleman RA, Abrams A, Marsh MP, Hryc CF, Ruths T, Chiu W, Ju T. Modeling protein structure at near atomic resolutions with Gorgon. J Struct Biol. 2011;174(2):360-73.

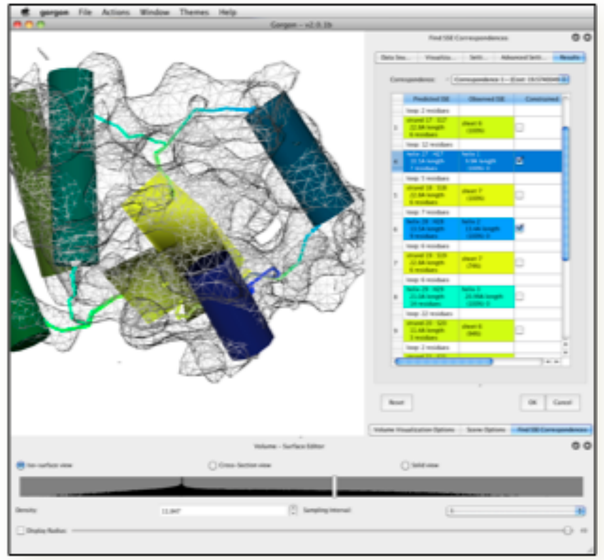
Model Building in Gorgon

- Density map and model visualization
- Density skeletonization (binary, grey-scale and interactive)
- SSE identification and building using SSEHunter
- SSE correspondence searches with helices and sheets
- Semi-automated atom placement
- Rigid body and flexible fitting



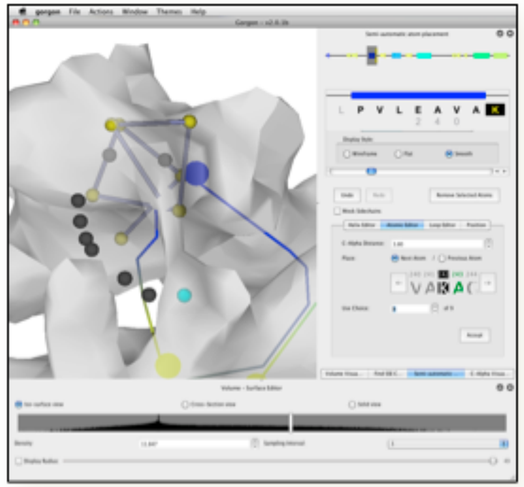


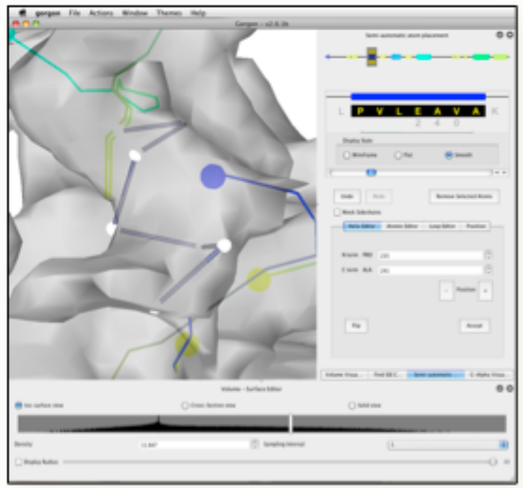
Modeling Tools

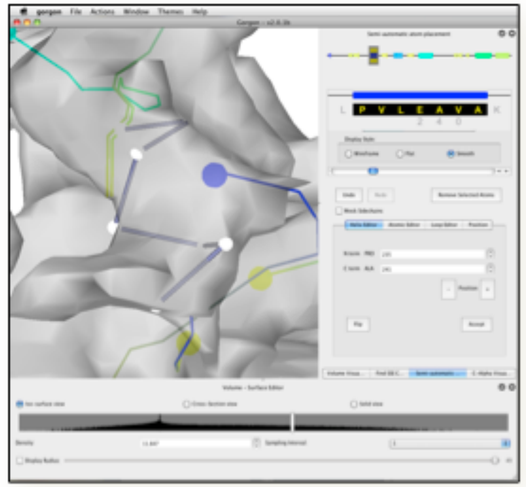


Sequence to structure correspondence using graph matching

- Based on helix position and distance/ connectivity
- Gallery of correspondences
- Partial assignments





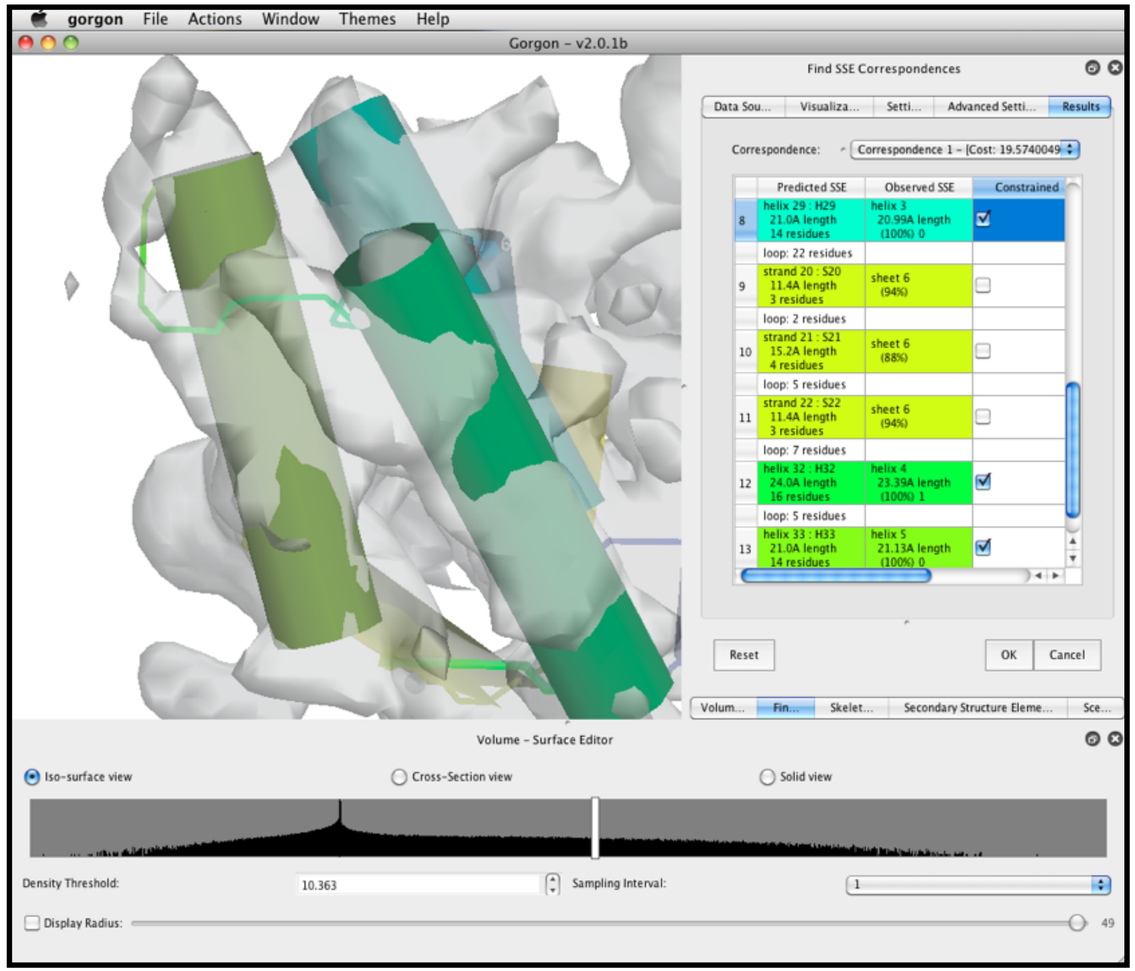


Interactive, semi-automated model building with density constraints

- Interactive, sketching of loops
- Auto-build of SSE
- Manual editing with local fitting

Features in Gorgon v2.2

- Improved SSE correspondence search with β sheets
- Complete SSEHunter and SSEBuilder integration
- Rigid-body and flexible fitting routines
- Session support to save and load work-in-progress
- Improved user-interface

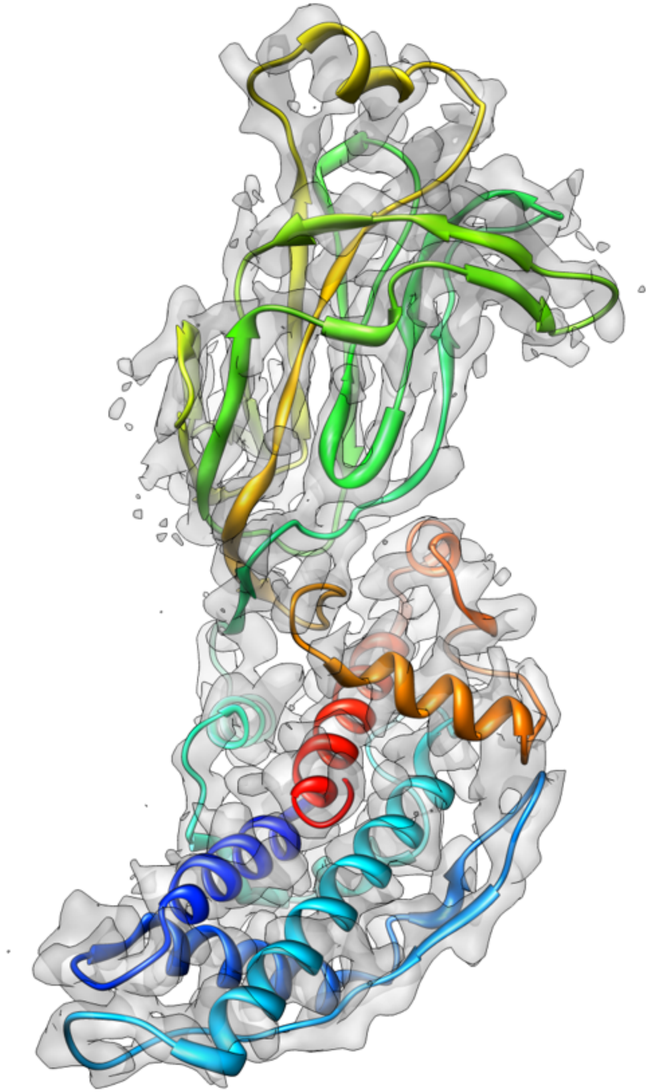


Hang Dou, Matthew Baker, Tao Ju. Graph-based deformable matching of 3D line segments with application in protein fitting. The Visual Computer (Proc. Computer Graphics International 2015), 31: 967-977.

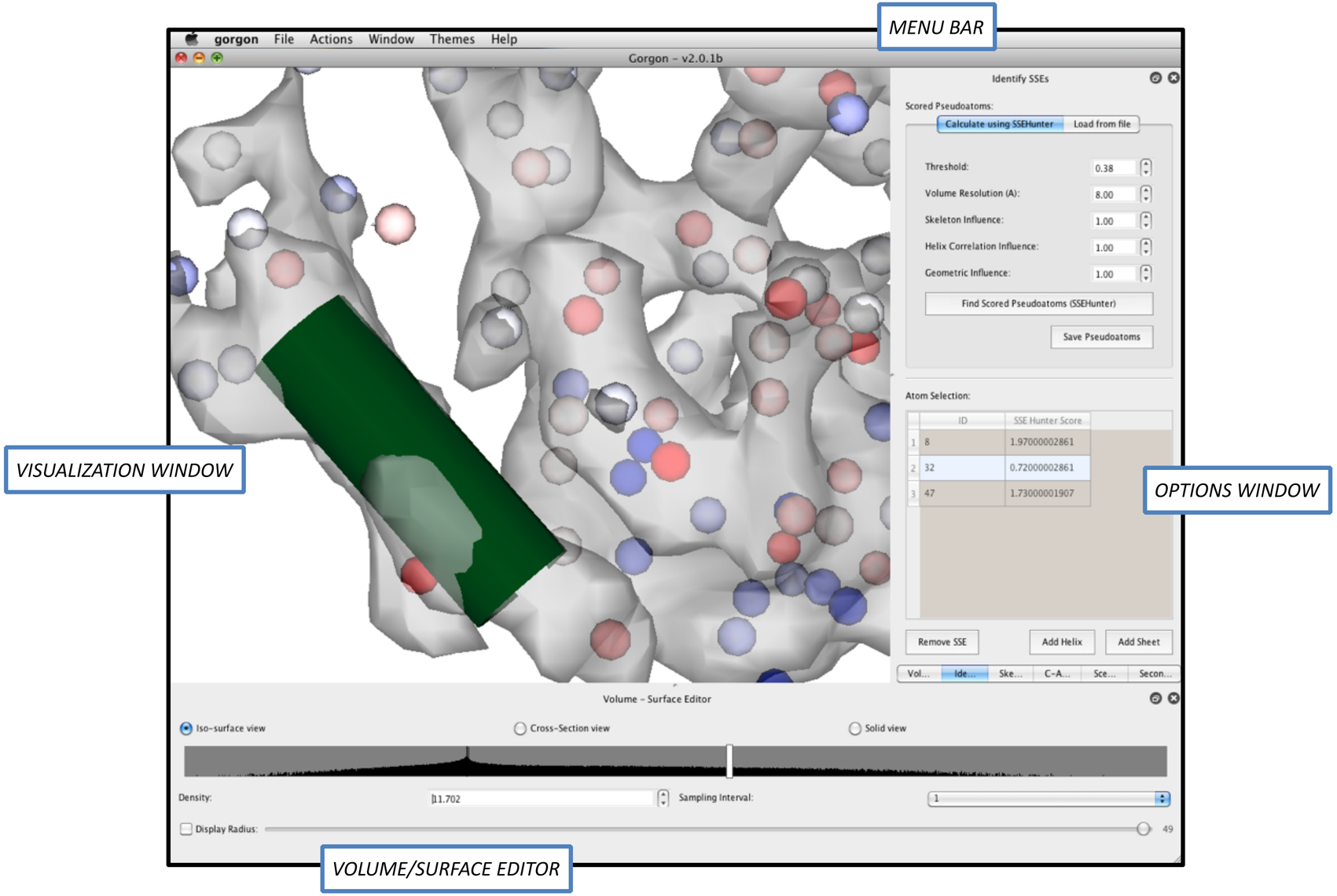
DE NOVO MODEL BUILDING TUTORIAL: GORGON

Model Building With Gorgon

- **The data:** Rotavirus VP6
  - 3.80 Å resolution
  - 1.23 Å/pixel
  - Monomer segmented with Chimera
  - EMDB ID: 1461
  - X-ray structure: 1QHD

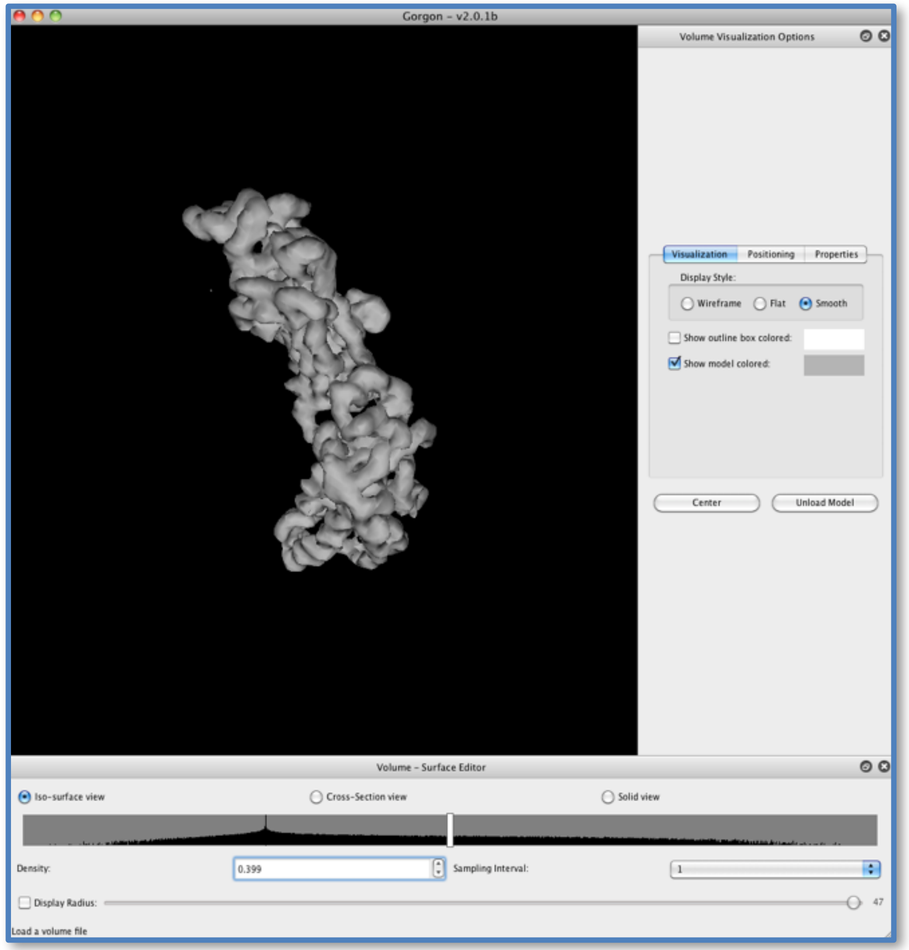


Gorgon Basics: Layout



Gorgon Basics: Opening a map

- File>Open>Volume "…Data-Sets/Gorgon/vp6-96o.mrc"
- Adjust transparency and color in options menu:
  - click on grey box next to "show model colored"
- Adjust isosurface in volume/surface editor options



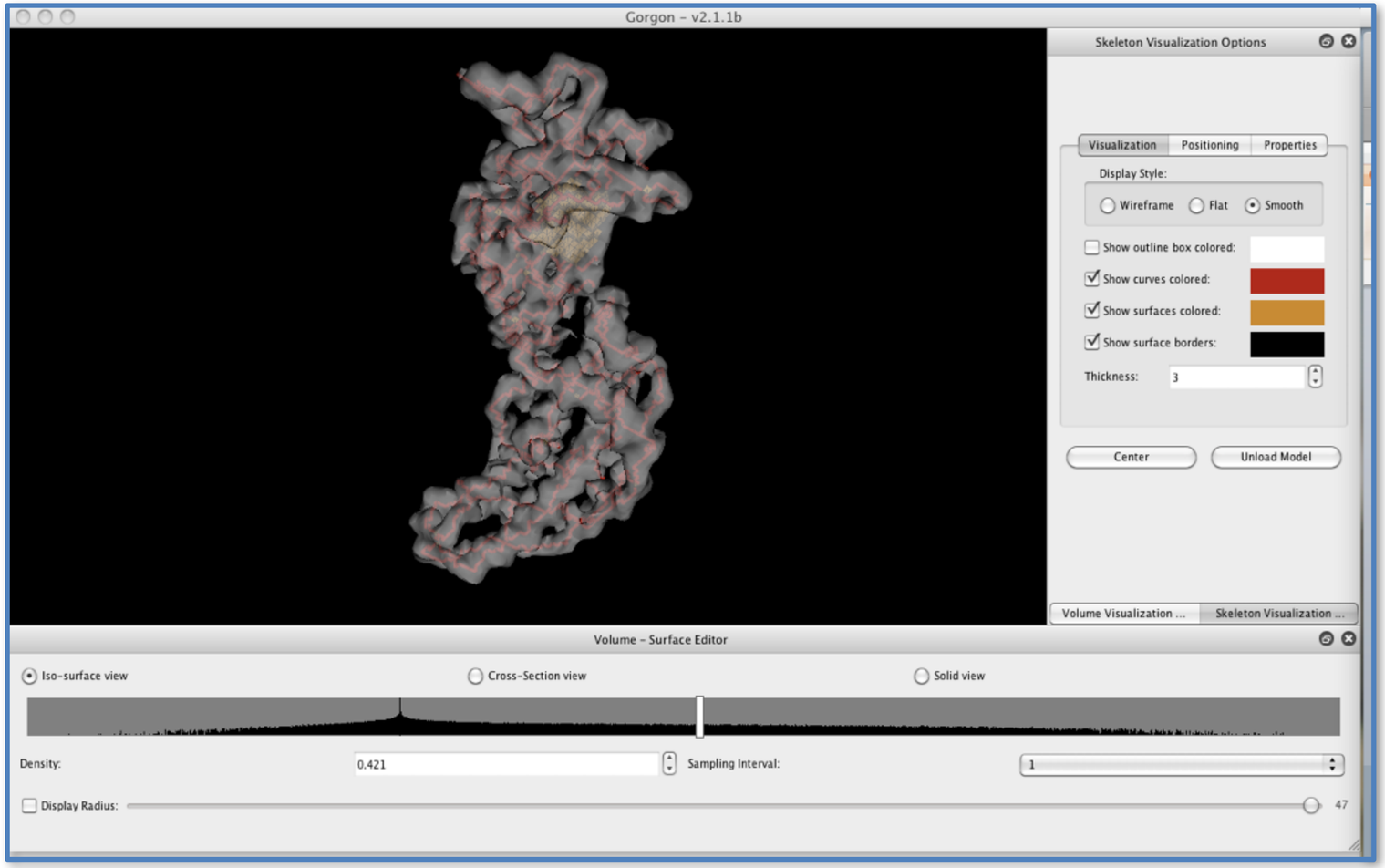
Gorgon Basics: Controls

	Left button	Middle button	Right button	Wheel
Click	***Selection***	***-***	***Focus (atom only)***	***Zoom***
Ctrl+click	***Toggle selection***	***-***	***Focus (atom only)***	***Change isosurface***
Click+drag	***Rotate***	***-***	***Translate***	***-***

Use the Apple key instead of Ctrl on a Mac

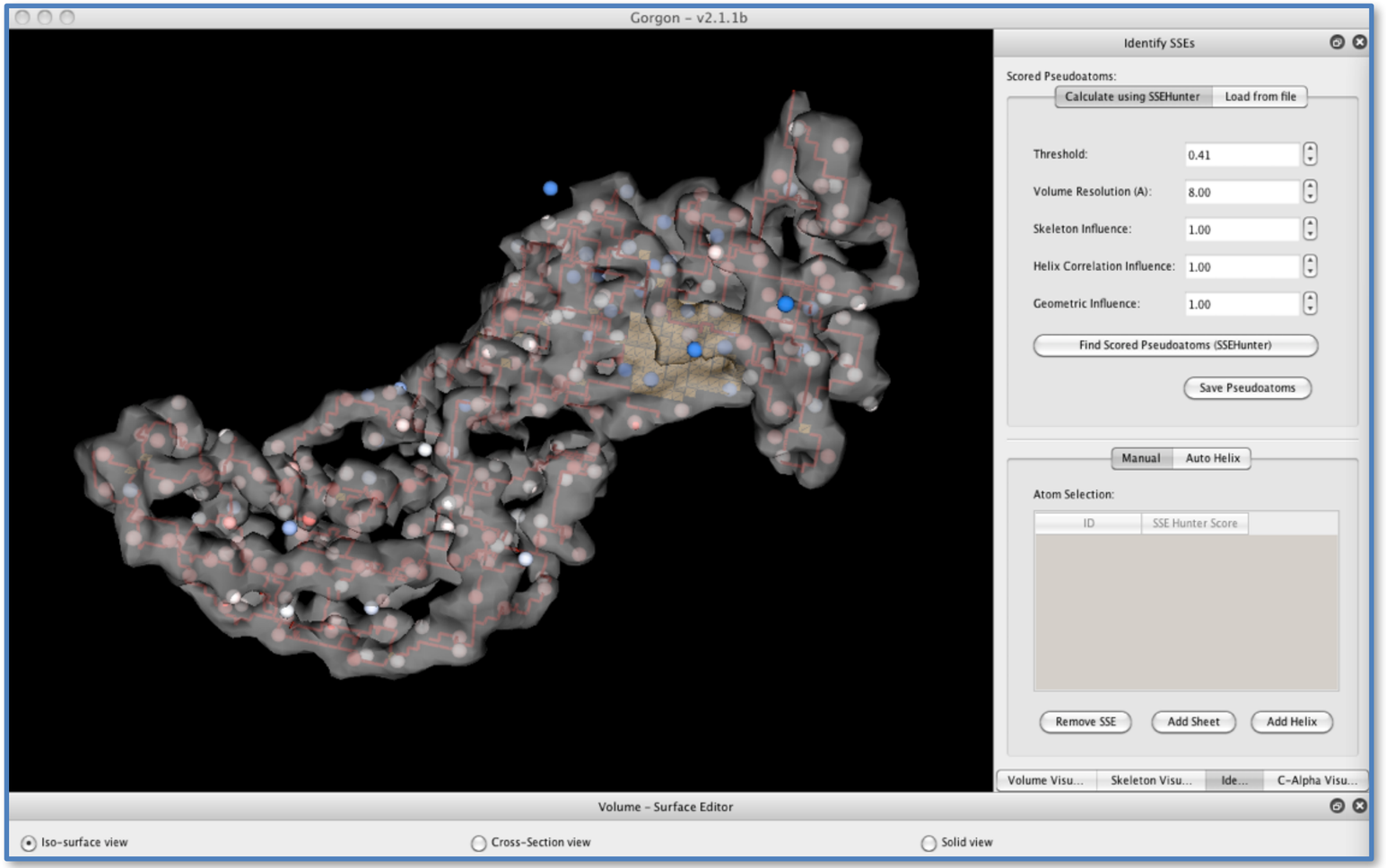
Gorgon: SSEHunter

- Build a skeleton
  - Actions> Volume> Skeletonization
  - Binary skeleton; select a threshold where the separation of stands and loops can be first seen (~0.40)
  - Sheets are in yellow, loops are in red



Gorgon: SSEHunter

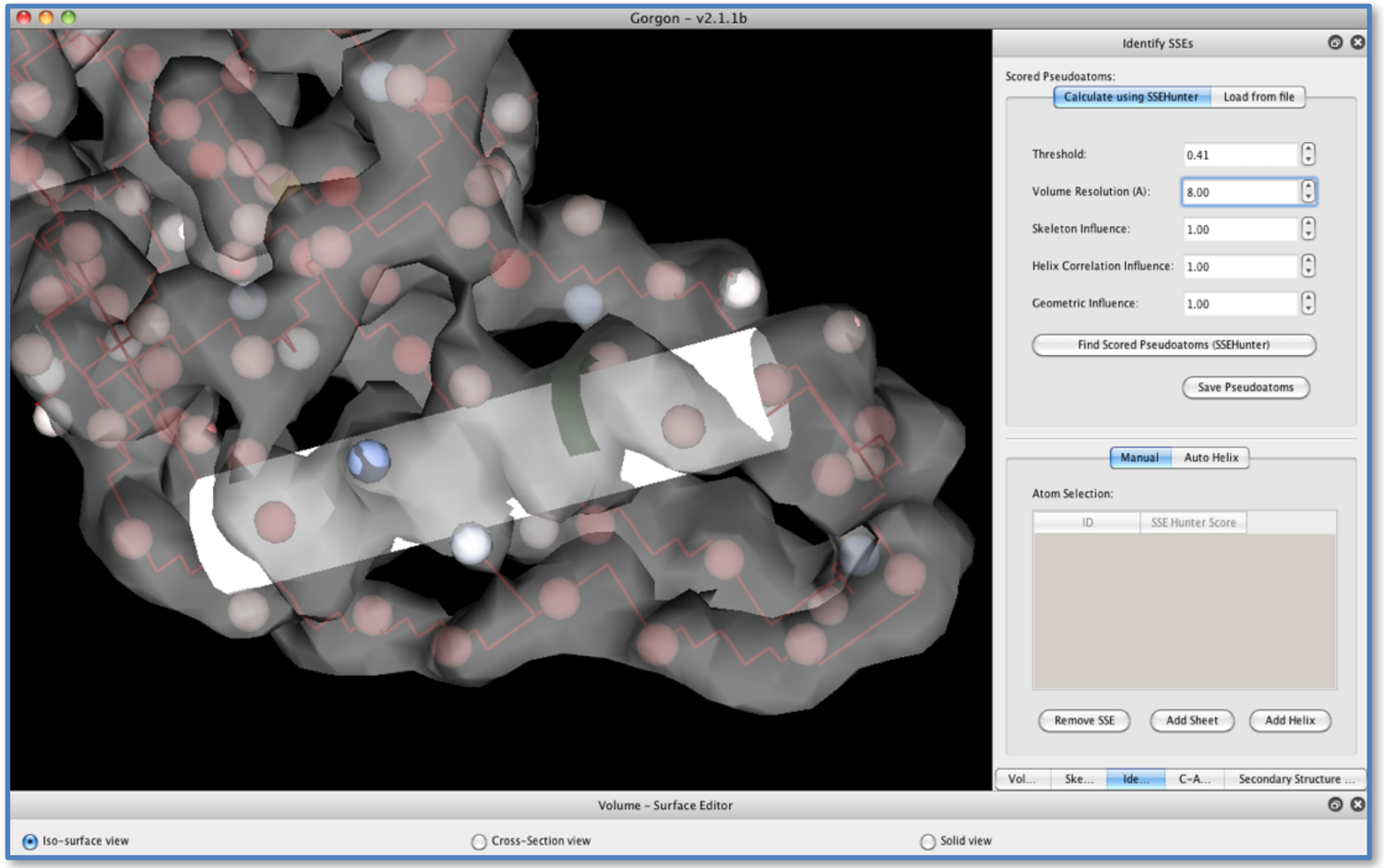
- Calculate SSEHunter scores
  - Actions> Secondary Structure Elements> Identify SSEs
  - Threshold ~0.40, resolution 8.00*
  - Click on Find Scored Pseudoatoms OR load from file “skeleton-vp6-b0.40.mrc”



* The resolution of the map is ~4Å; it is not necessary to adjust this parameter in general when working with high resolution data

Gorgon: SSEHunter

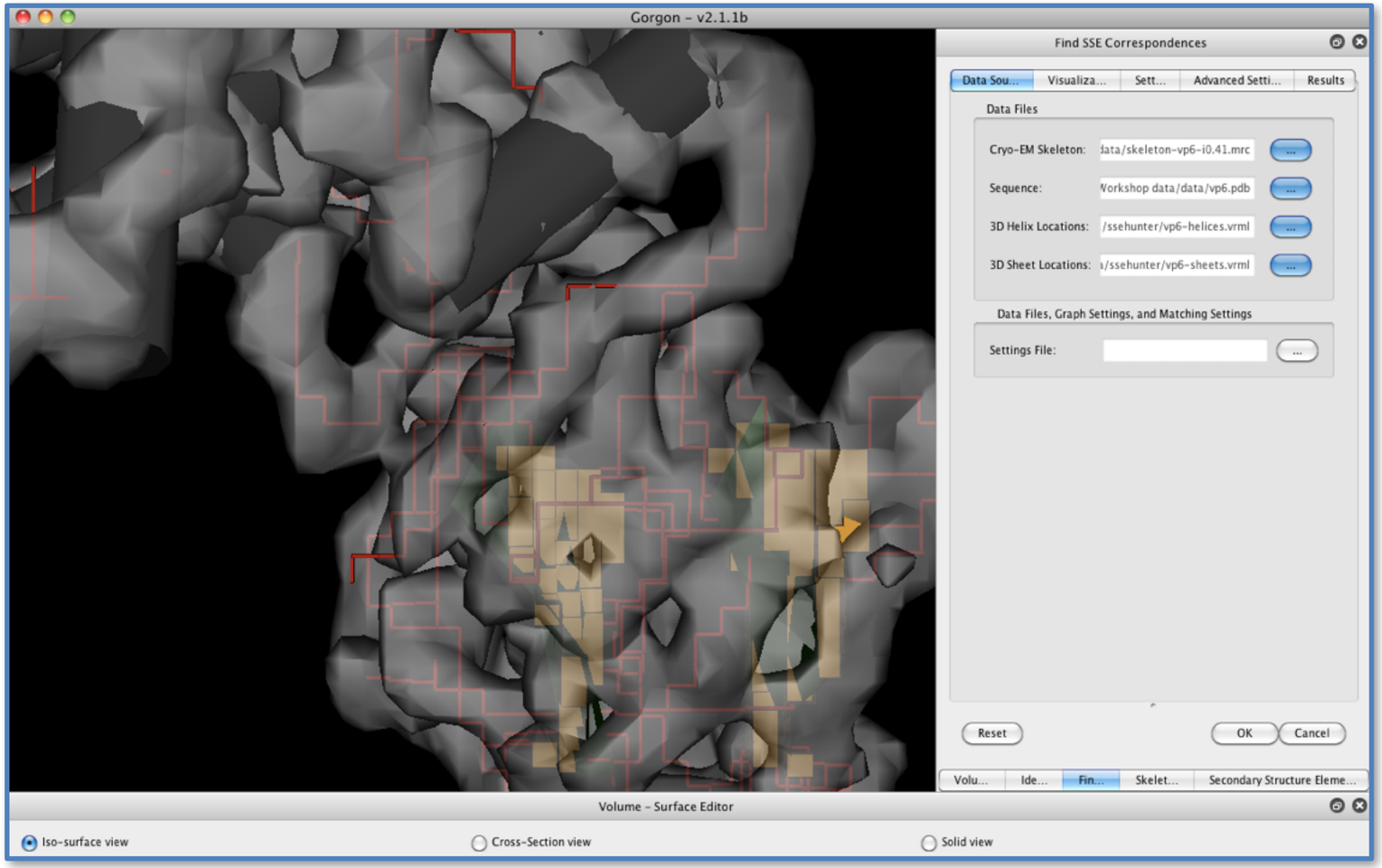
- Build SSE models
  - Ctrl+click on red Ca-atoms at beginning and end of helix
  - Click “Add Helix”
  - Select Helix in visualization window and press Ctrl+f to refine fit to density
  - Ctrl+click on blue Ca-atoms in a plane
  - Click “Add Sheet”
  - Iterate until all visible SSEs are annotated



*** When finished, close Ca file (the ssehunter score) in the file menu and save helices/sheets as VRML ***

Gorgon: SSE Correspondence

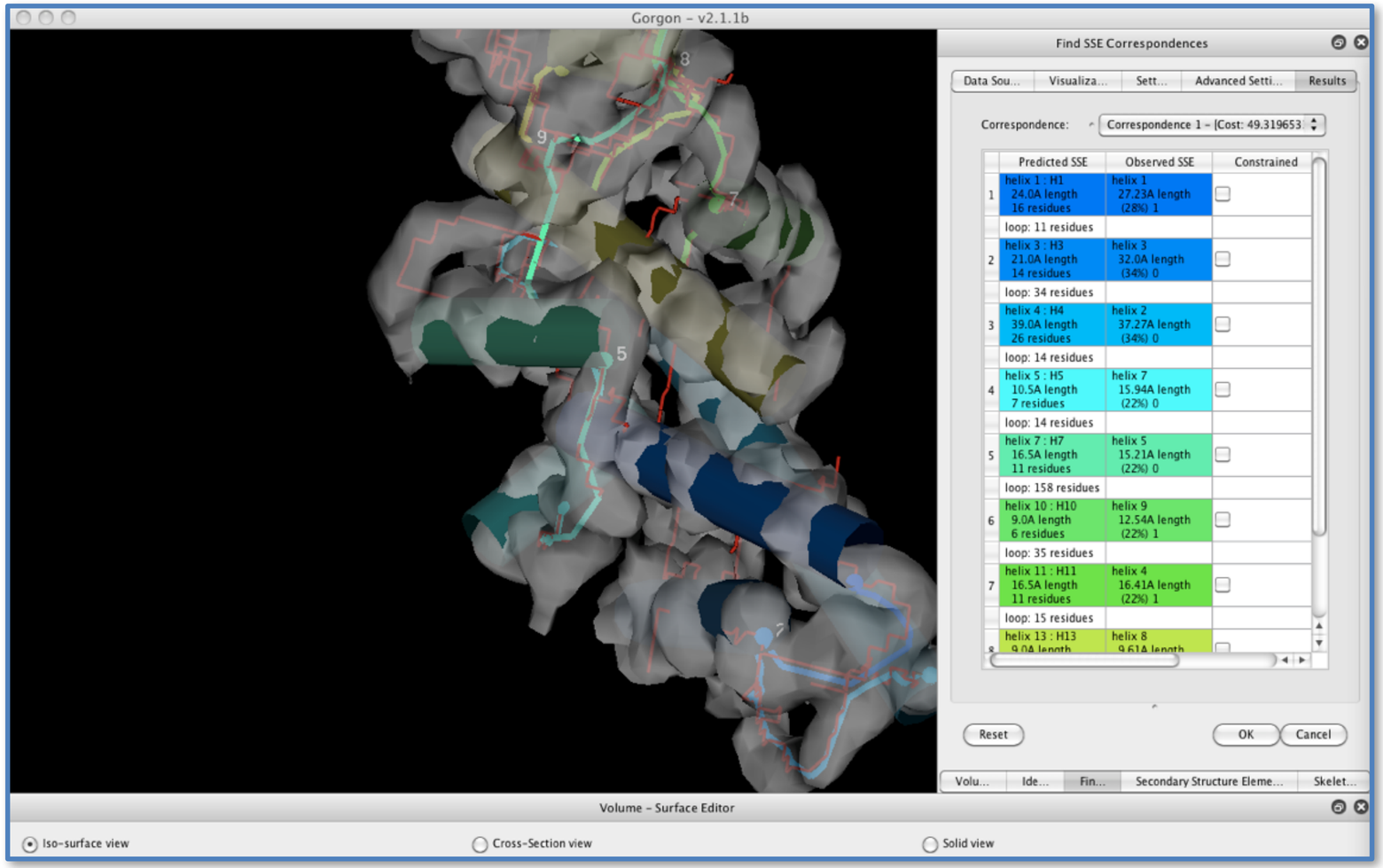
- Generating an SSE correspondence
  - Actions > Secondary Structure Elements > Find SSE correspondence
- Files
  - Cryo-EM skeleton: “skeleton-vp6-b0.40.mrc”
  - Sequence: “vp6.pdb”
  - 3D Helix locations: “helices-vp6.vrml”
  - 3D Sheet locations: “sheets-vp6.vrml”
- Click “OK”



*** Close all SSEs (VRML files) from the previous step before loading the SSEs in the example ***

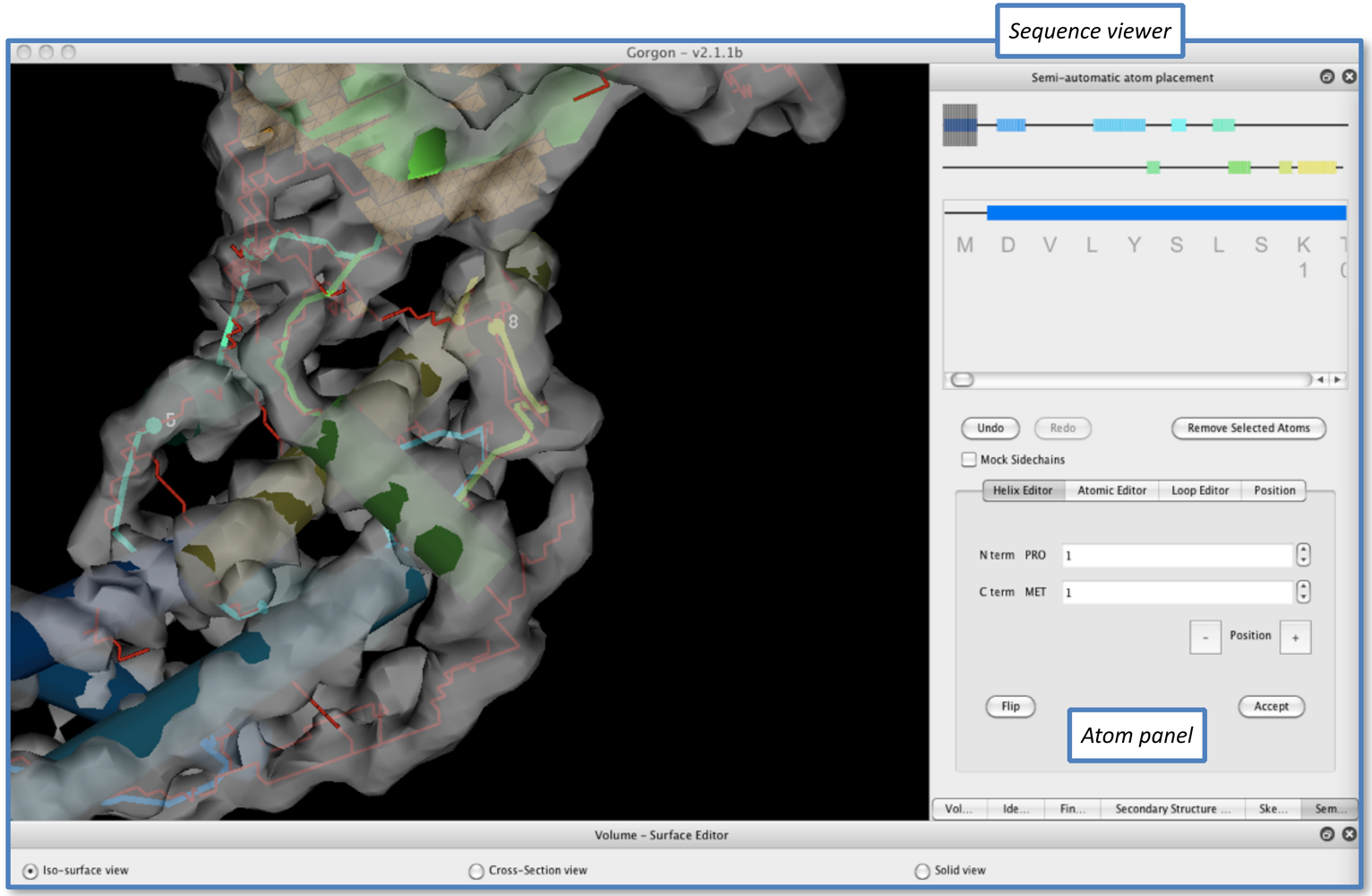
Gorgon: SSE Correspondence

- Select a correspondence
  - Evaluate correspondence by comparing lengths, percentiles and overall topology
  - Click on helices in visualization window to view correspondence
  - Constrain “good” matches by clicking on “Constrained” box
  - Click “OK” to re-run correspondence routine with desired selection



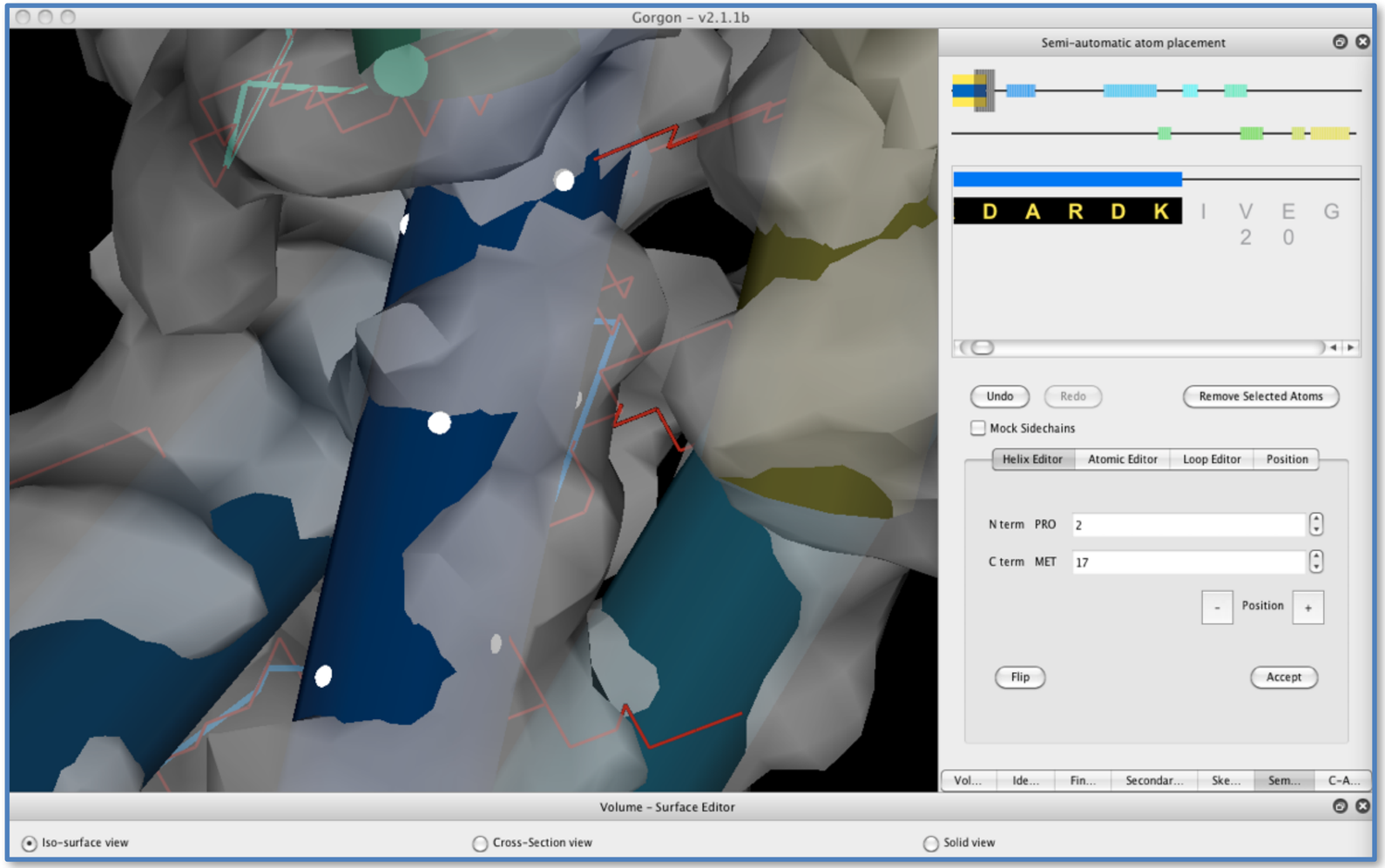
Gorgon: Model Building

- Atom placement
  - Actions > C-alpha atoms > Semi-automated atom placement
  - Top panel is sequence viewer with SSE predictions
  - Current location in sequence is shown in grey bar and below in the zoom view
  - Bottom panel (atom panel) has 4 tabs for atom placement



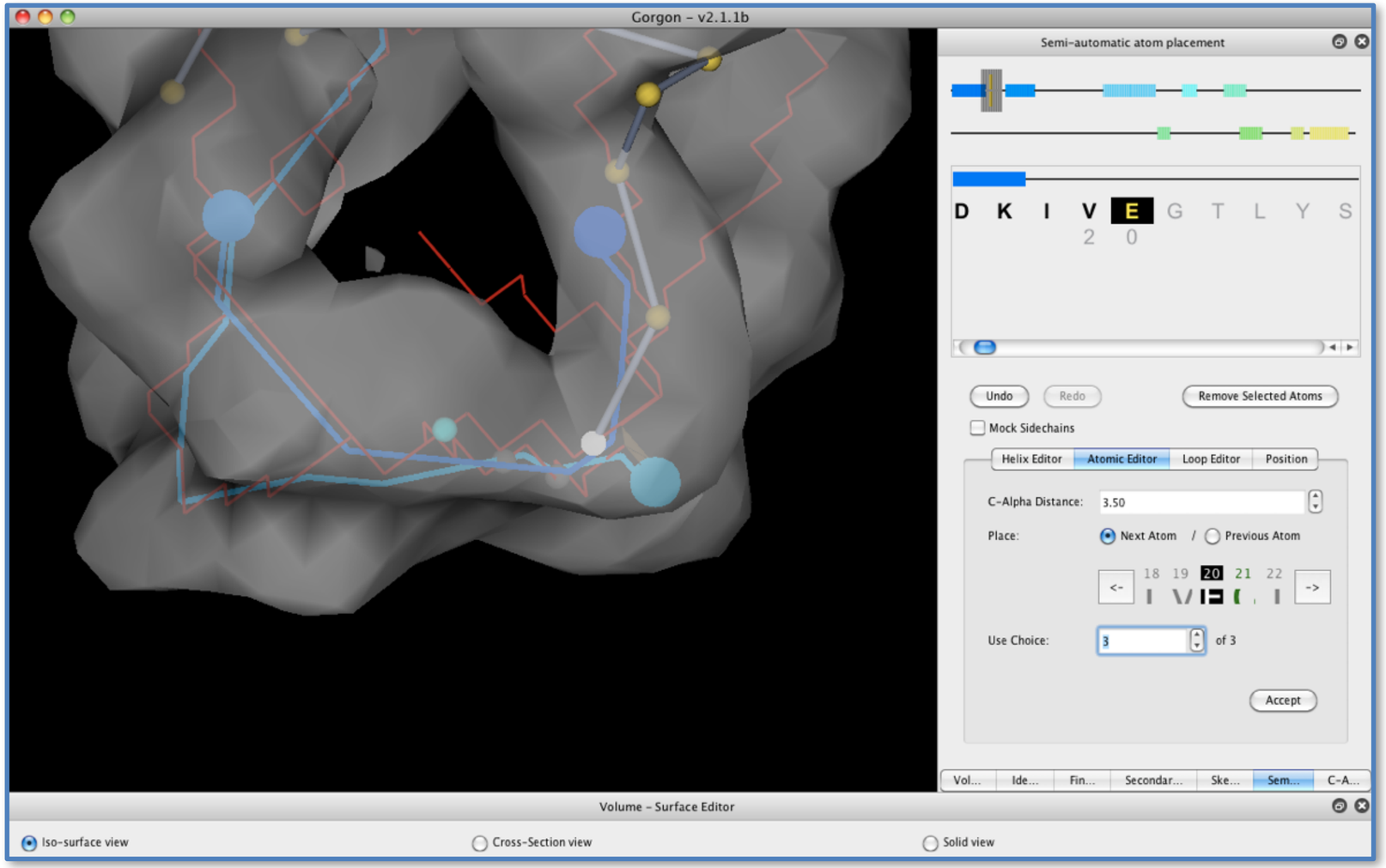
Gorgon: Model Building

- Add a helix
  - Select “helix editor” in the atom panel
  - Ctrl+Click on a helix in the visualization window
  - Click on the corresponding helix segment in sequence viewer; sequence will be highlighted in black
  - Adjust length/position in the helix editor if desired
  - Click “Accept” to build a helix; atoms will appear in visualization window
  - Click on an assigned atom an locate it in the sequence view
  - If helix is reversed, flip helix by clicking on “flip” (at least 1 atom in the helix must be selected
  - Repeat until all helices are assigned



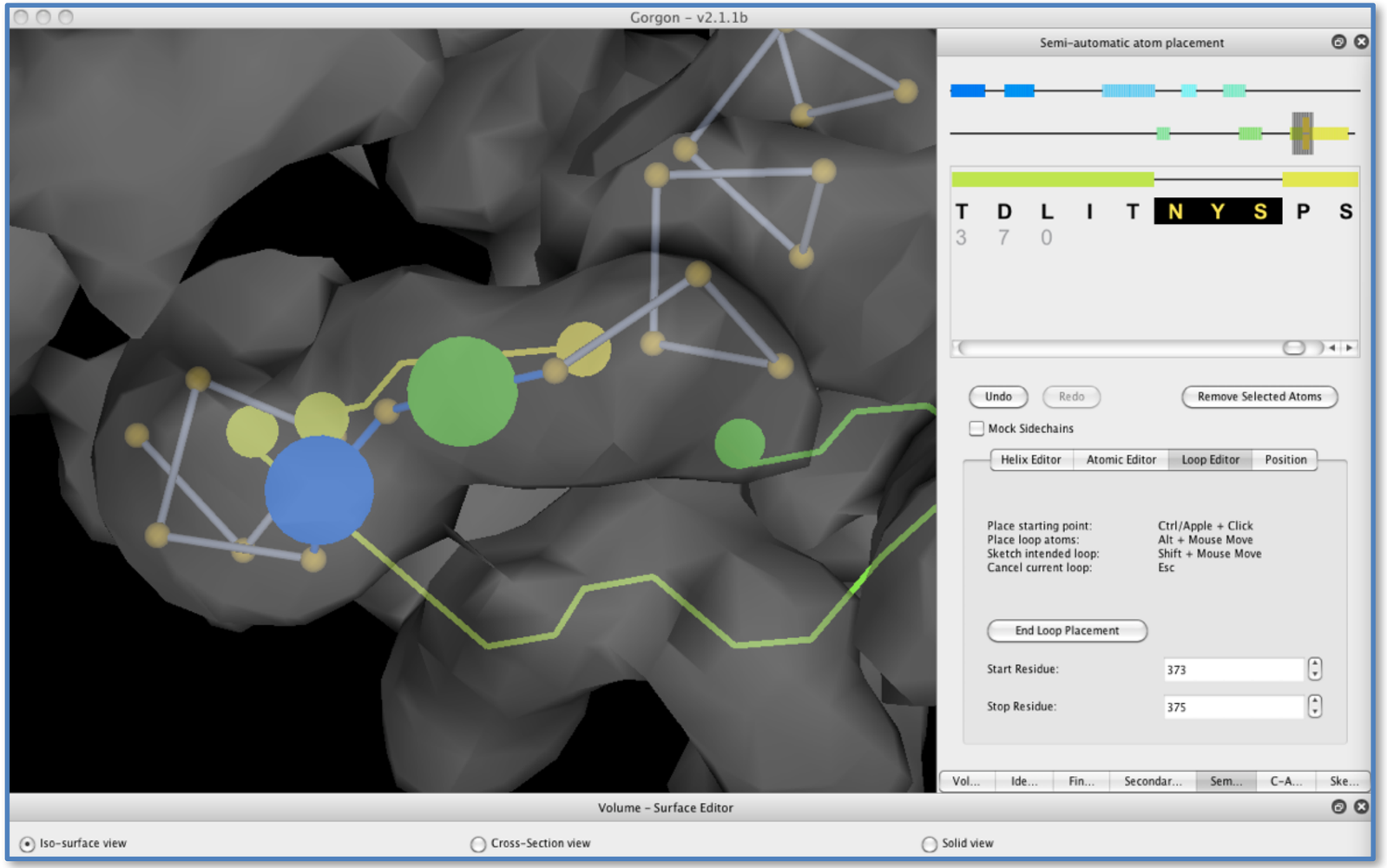
Gorgon: Model Building

- Build a loop
  - Select “atomic editor” in the atom panel
  - Set C-alpha distance to 3.5Å
  - Click on a starting/ending atom in the vis window
  - Select direction in Atomic editor panel (next atom increments, previous atom decrements)
  - Selected residue is highlighted, residue to be placed is in green
  - Cycle through the “Use choice” positions to find best placement, current position for the atom to be placed is in cyan
  - Click on “Accept”
  - Repeat until next assigned atom or terminus is reached



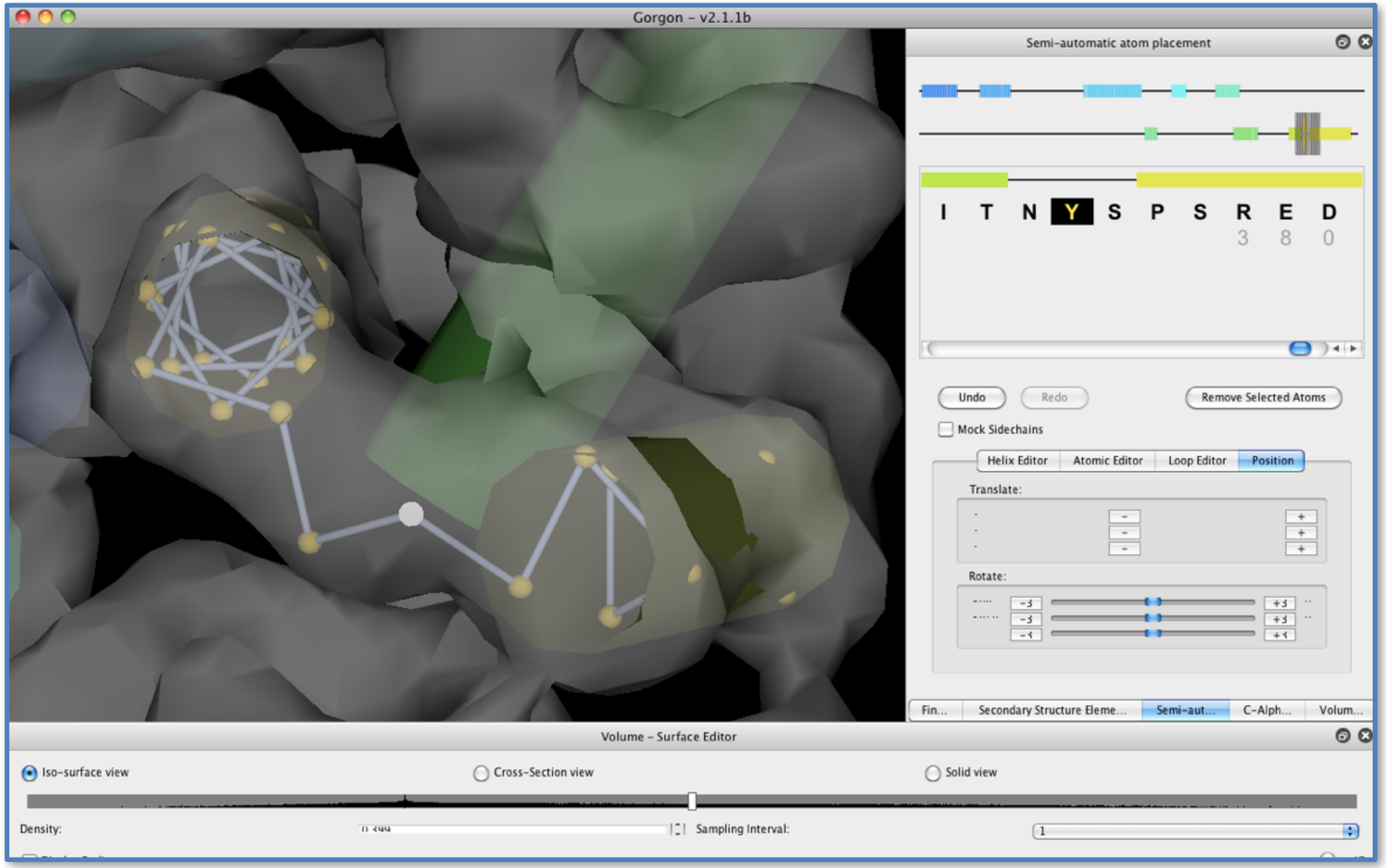
Gorgon: Model Building

- Adding a loop
  - Select “loop editor” in the atom panel
  - Select residues between two assigned residues in the sequence window
  - Click “start loop placement”
  - Select start point by Ctrl+ click on the desired start point
  - Move loop through density with alt+move
  - Click “End loop placement” when finished



Gorgon: Model Building

- Adjust atom positions
  - Select residue (click)
  - Adjust position by Ctrl+click +drag or use Position editor in Atom panel
    - Blue bonds are too short(<3.5Å)
    - Red bonds are too long (>4.2Å)
    - Relative sidechain size can be shown by selecting “mock sidechains”
  - Repeat until all atoms are adjusted



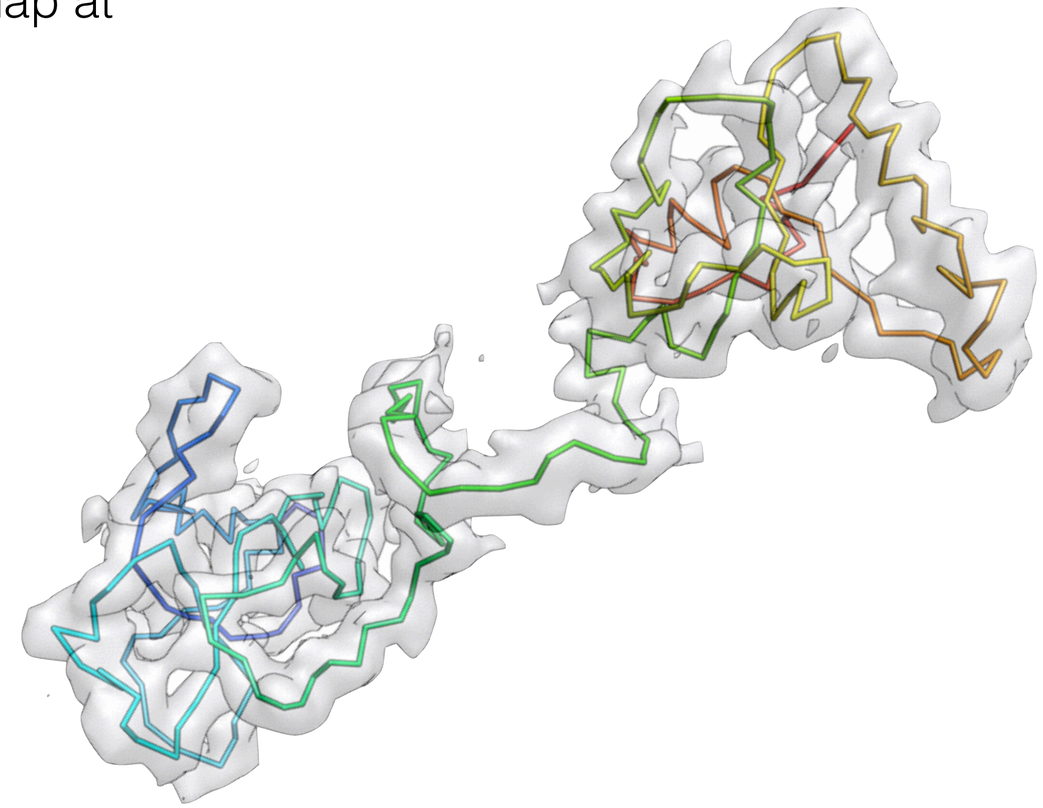
*** When finished, save model in the File menu as C-alpha atoms or use export to PDB***

DE NOVO MODEL BUILDING TUTORIAL: PATHWALKING

Deriving a Model With Limited Constraints

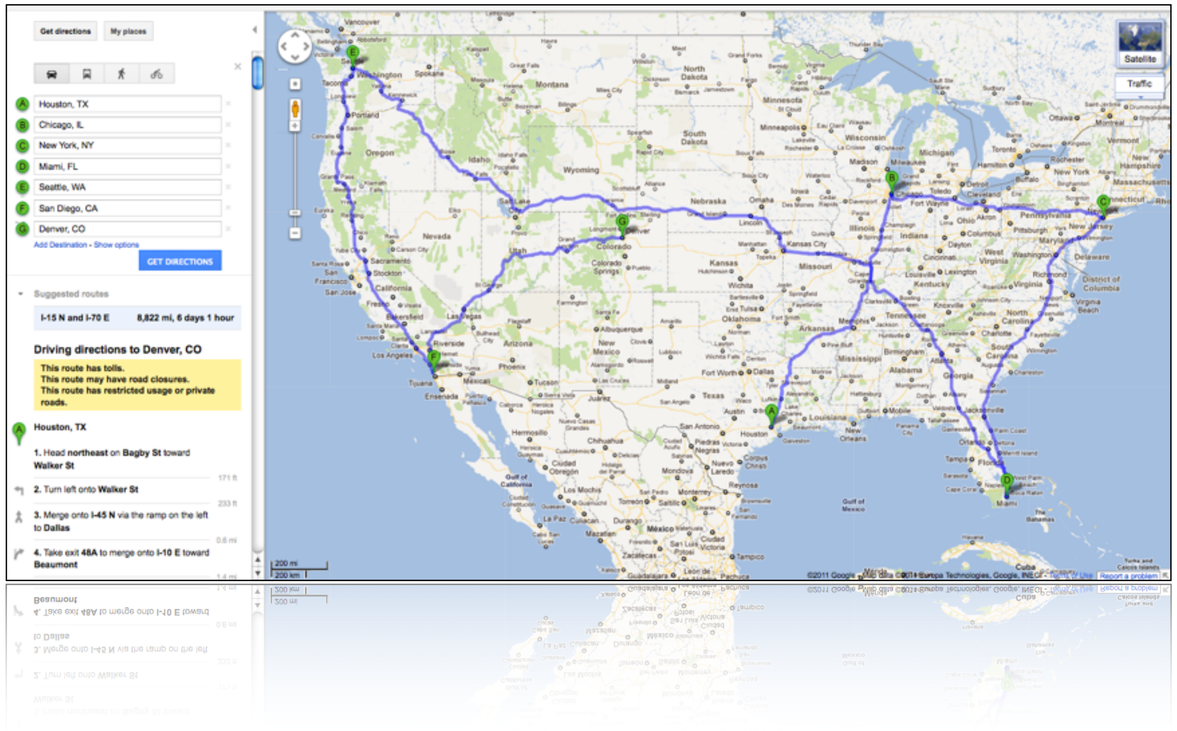
Goal: Find a path or sets of paths that trace the complete path of a protein through a density map at near-atomic resolutions such that:

- ➡ No SSEs required
- ➡ No explicit sequence information required
- ➡ No structural template required
- ➡ Automated
- ➡ Optimized against biophysical constraints

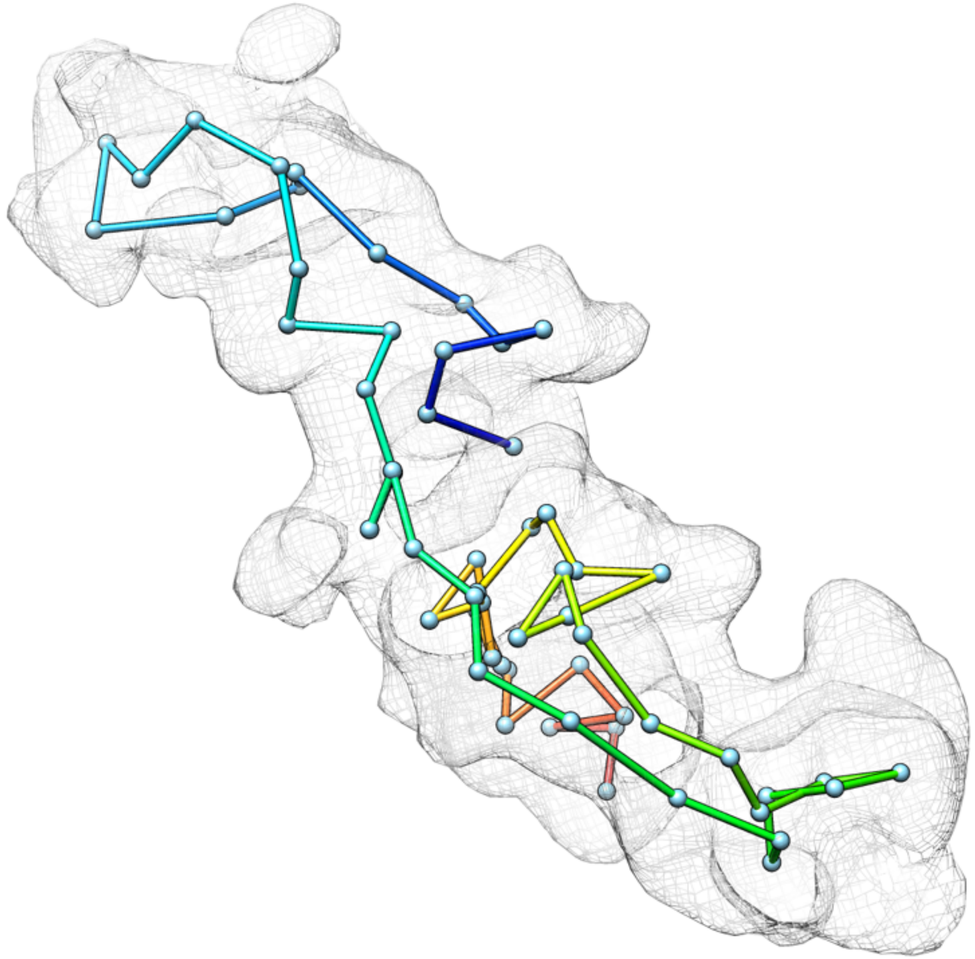


Repurposing the Traveling Salesman Problem

- TSP calculates optimal route between cities by minimizing distance travelled
- Each city can only be visited once
- Several exact and approximate TSP solvers for thousands of nodes



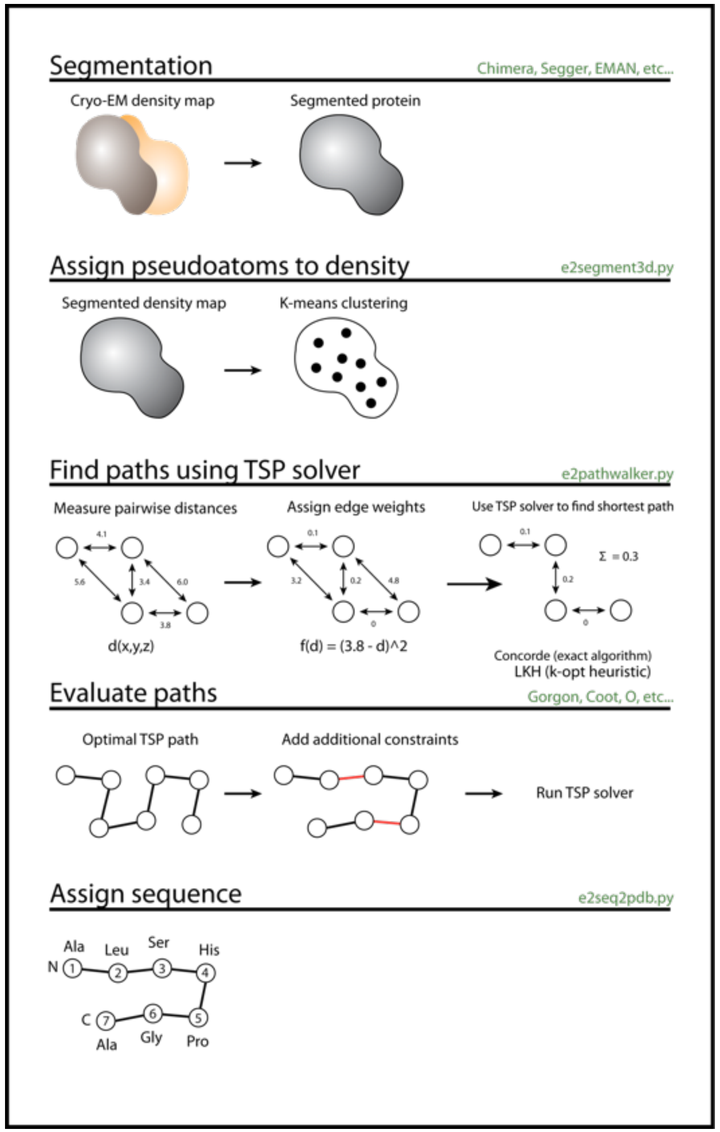
Connecting the “Dots”



Protein Structure Determination with TSP

Re-pose de novo modeling as a 3D TSP problem

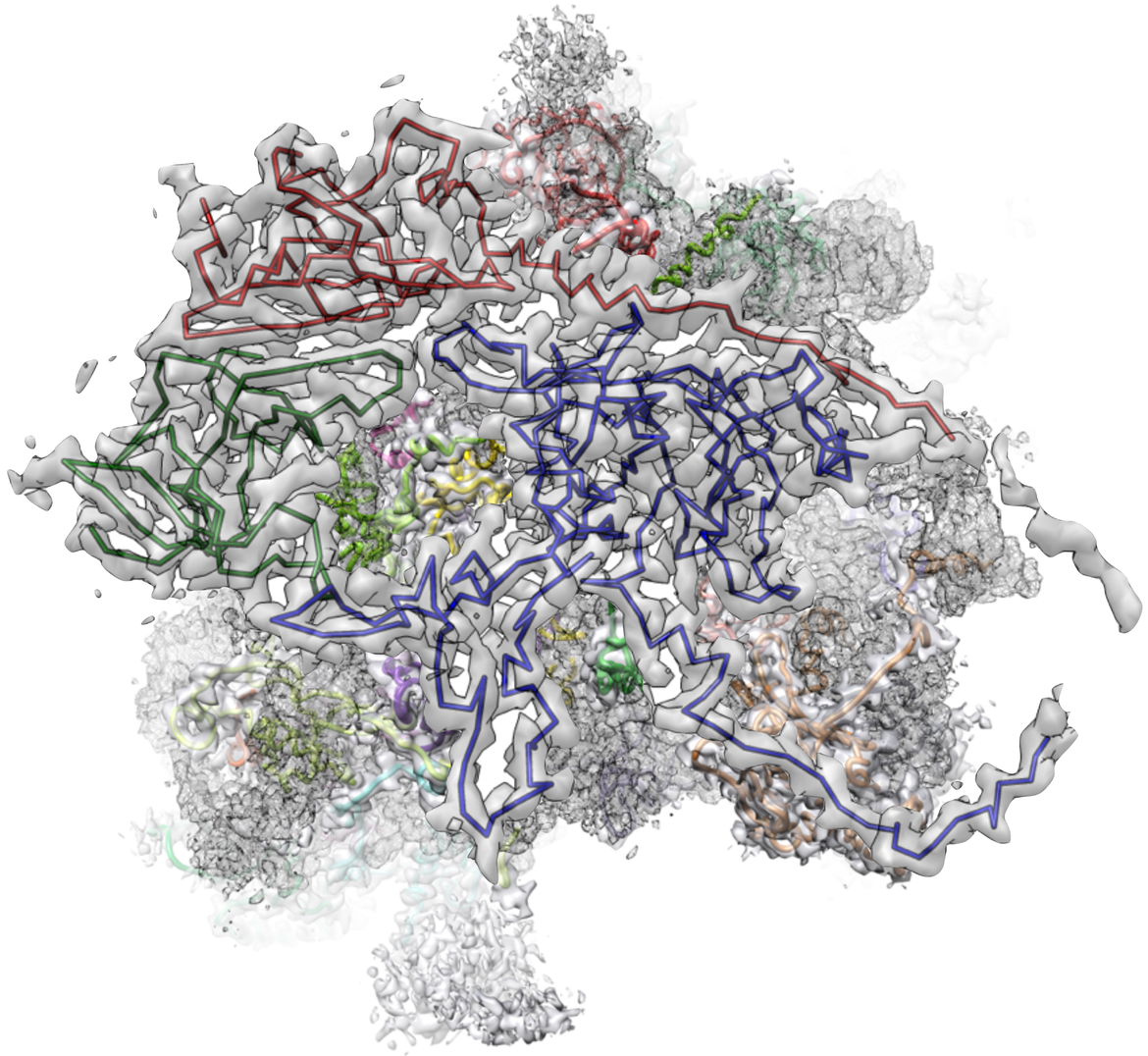
- ➡ C-alpha atoms are “cities”
- ➡ Protein backbone is not a minimal distance, rather optimal distance is 3.8Å*N
- ➡ Distance expressed as a deviation from 3.8Å



Baker MR, Rees I, Ludtke SJ, Chiu W, Baker ML. Constructing and validating initial Cα models from subnanometer resolution density maps with pathwalking. Structure. 2012 Mar 7;20(3):450-63.

Building a Better Pathwalker

- Optimized pseudoatom generation
- Fully automated with “path checking"
- Density weighted paths
- Iterative SSE detection and pseudoatom placement
- Geometry filtering
- Sidechain filtering
- Multiple chains



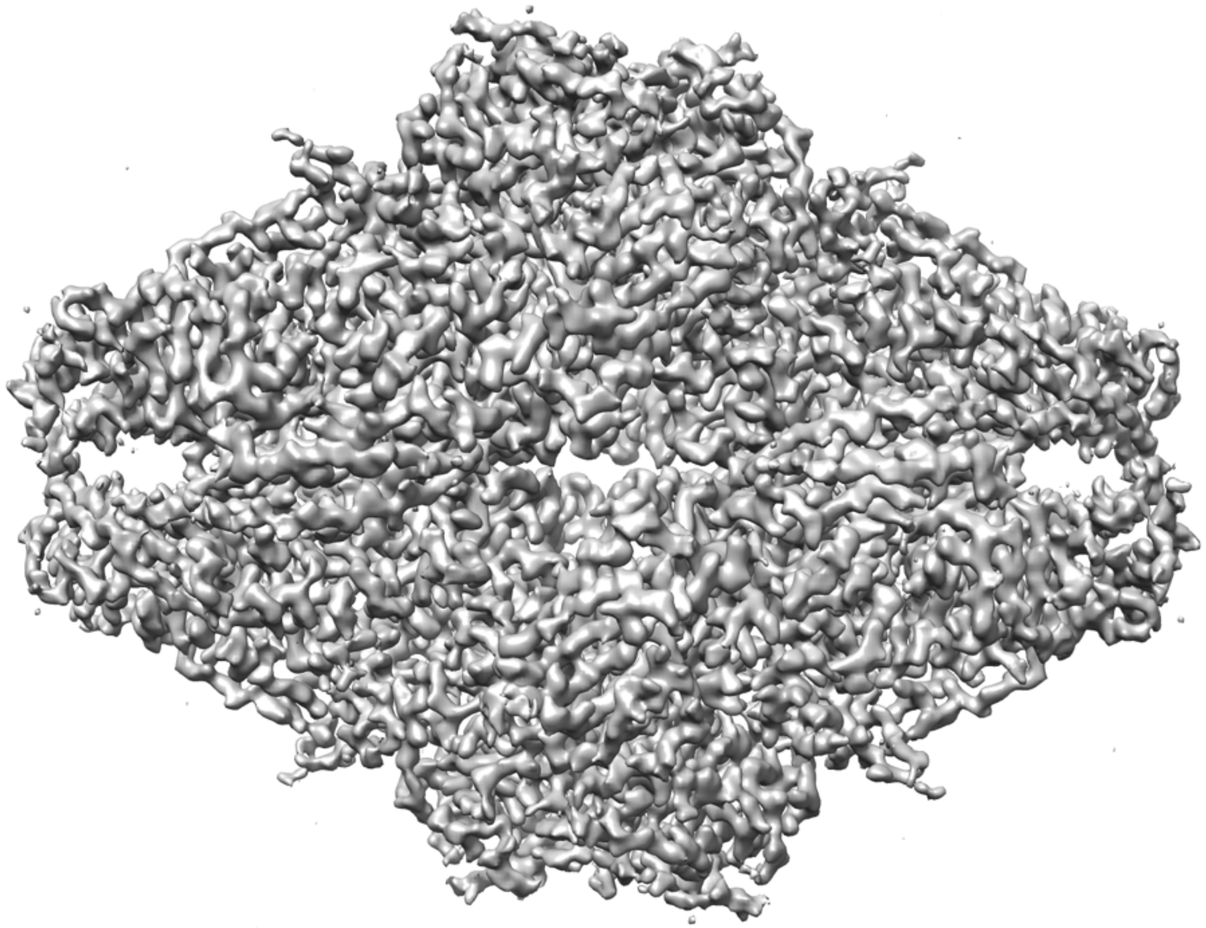
Software

- EMAN2: cryo-EM image processing; pathwalker, a sequence free modeling tool
  - http://blake.bcm.tmc.edu/eman/eman2/
- UCSF Chimera: Visualization
  - http://www.cgl.ucsf.edu/chimera/

All software available for windows, OS X and linux

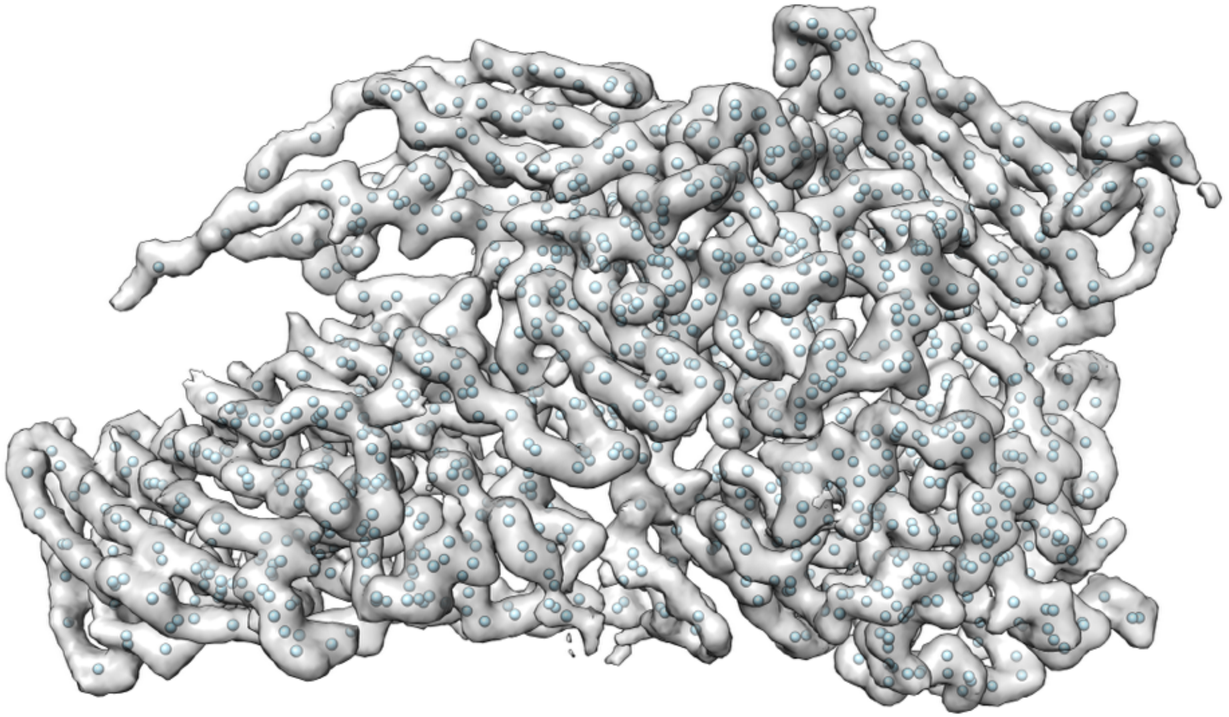
Model Building With Pathwalker

- **The data**: Beta-Galactosidase
  - 3.2 Å resolution
  - 0.6375 Å/pixel
  - Monomer segmented with Chimera
  - EMDB ID: 5995
  - X-ray structure: 3j7h



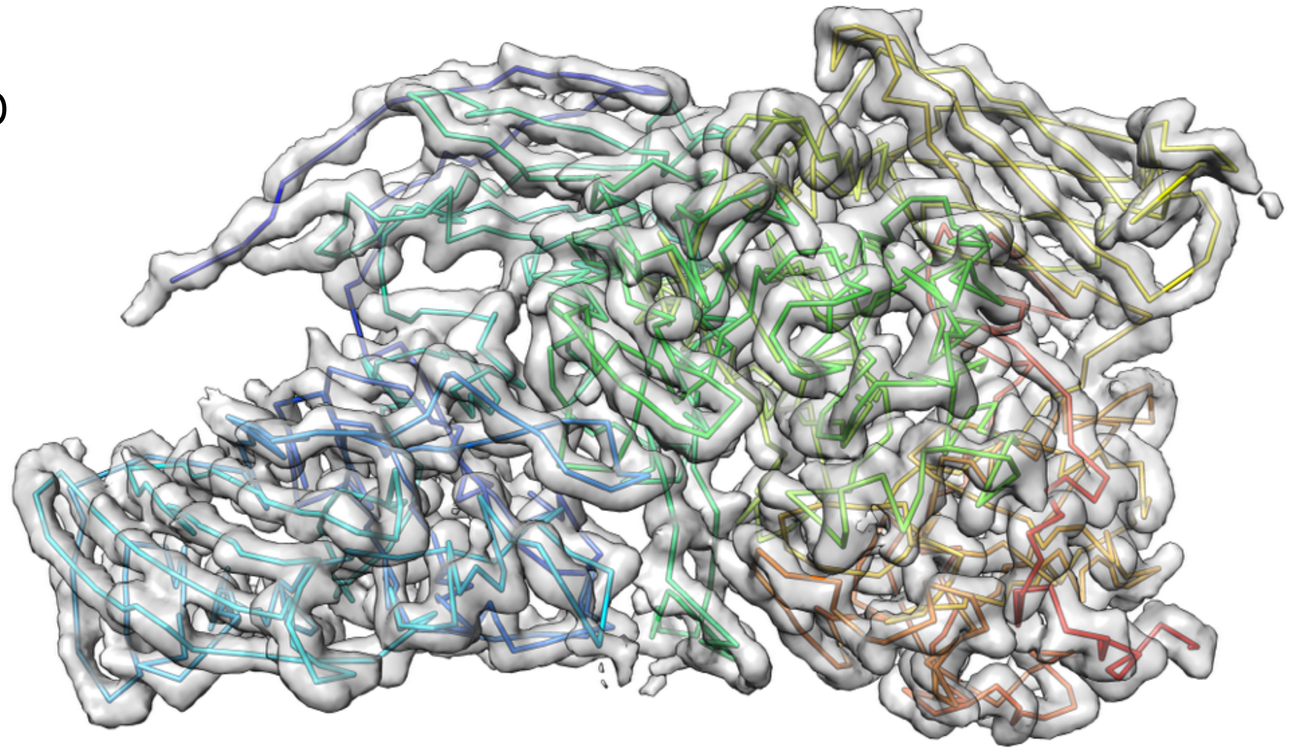
Pathwalker: Generating pseudoatoms

- In a terminal window run
  - e2proc3d.py sub-A.mrc map.mrc --process normalize.edgemean --process threshold.belowtozero
  - e2segment3d.py map.mrc --pdbout=pseudoatoms.pdb --process=segment.kmeans:ampweight=1:nseg=1022:verbose=1:minsegsep=1:pseudoatom=1:thr=10
- Open pseudoatoms.pdb file in Chimera
  - Show only atoms, not bonds or ribbons



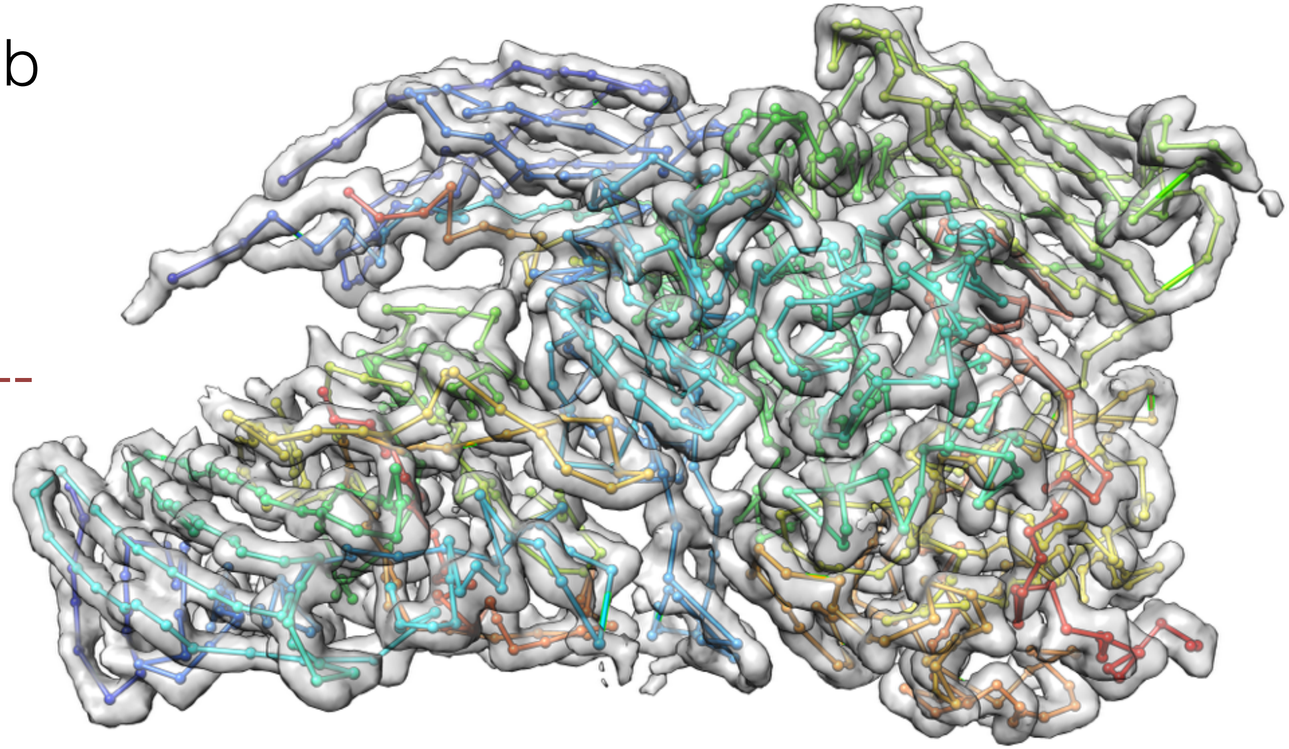
Pathwalker: Calculating an initial path

- In a terminal window run
  - e2pathwalker.py pseudoatoms.pdb --mapfile=map.mrc --output=path0.pdb --solver=lkh --overwrite --dmin=1 --dmax=10 --mapthresh=12 —mapweight=200
  - Open path0.pdb file in chimera
  - Render as ball and stick



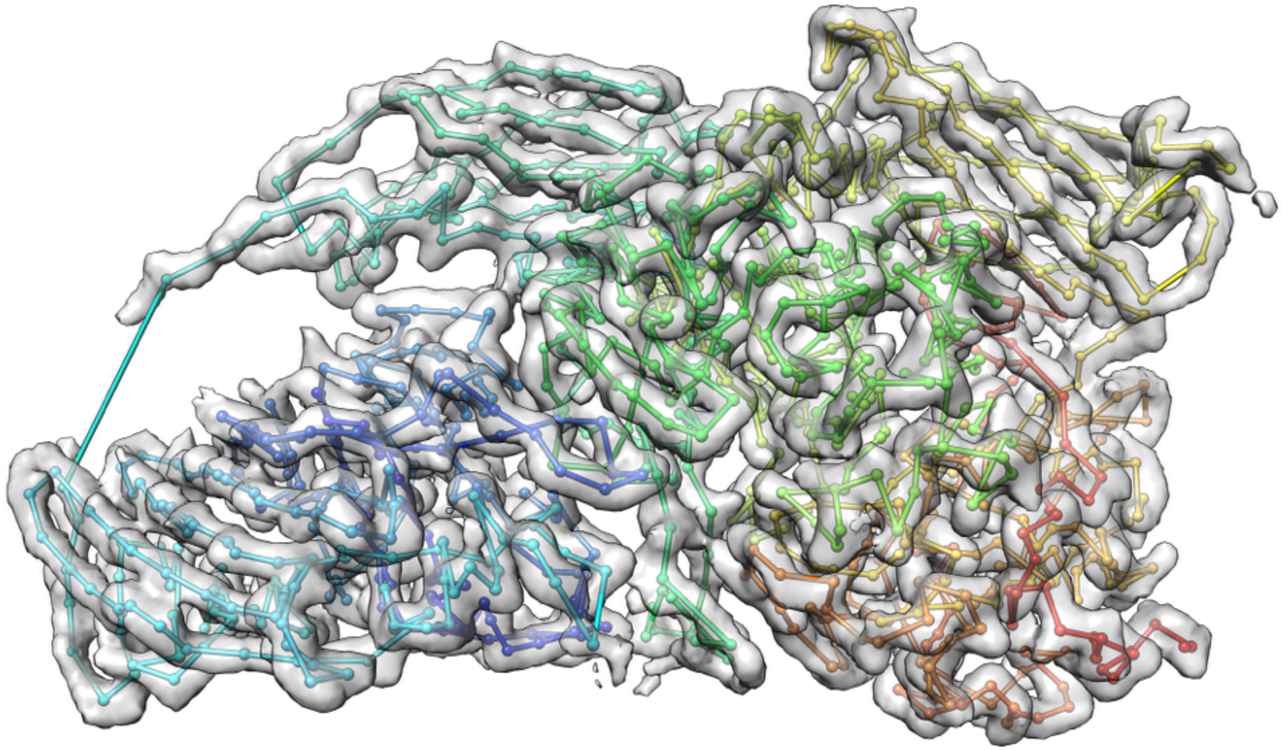
Pathwalker: Re-calculating the path

- In a terminal window run
  - e2pathwalker.py pseudoatoms.pdb --mapfile=map.mrc --output=path1.pdb --solver=lkh --overwrite --dmin=1 --dmax=10 --mapthresh=12 --mapweight=200 **--subunit=3**
  - Open path1.pdb file in chimera
  - Render as ball and stick



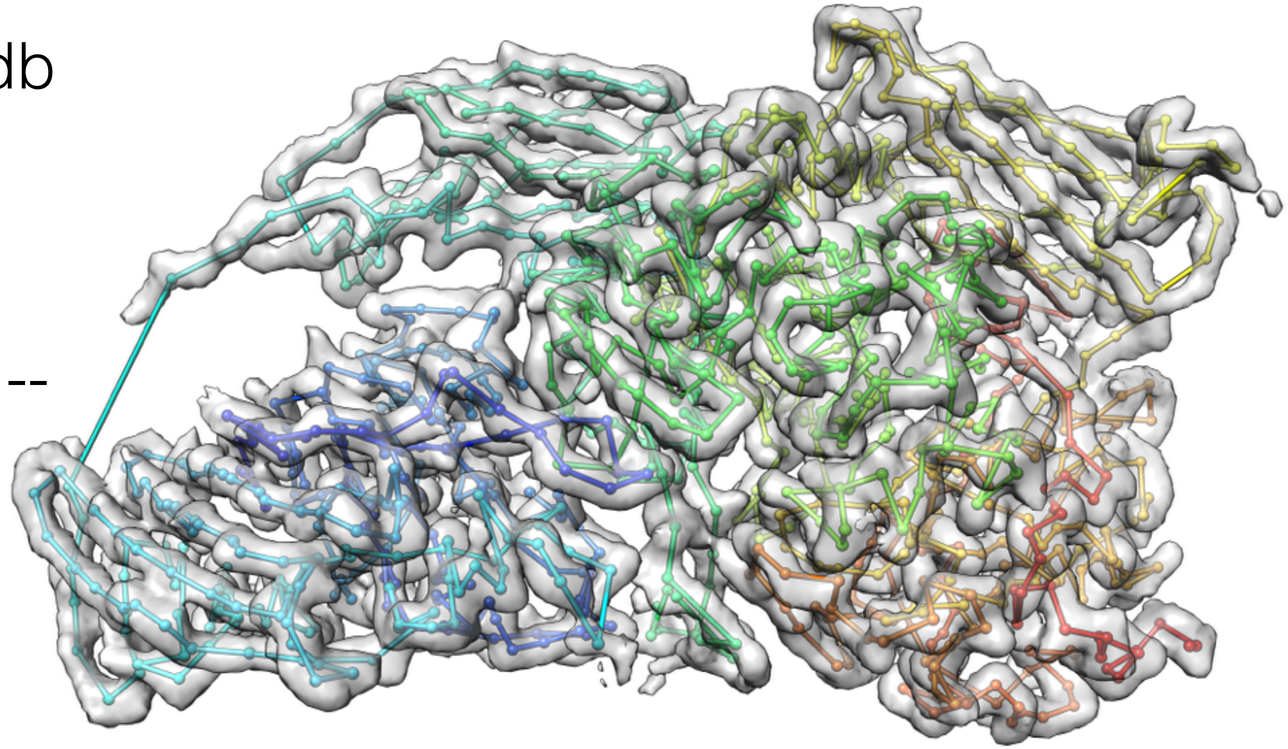
Pathwalker: Fixing edges in the path

- In a terminal window run
  - printf "870 1017\n827 829\n" > edge.txt
  - e2pathwalker.py pseudoatoms.pdb --mapfile=map.mrc --output=path2.pdb --solver=lkh --overwrite --dmin=1 --dmax=10 --mapthresh=12 --mapweight=200 --**edgefile=edge.txt**
  - Open path2.pdb file in chimera
  - Render as ball and stick



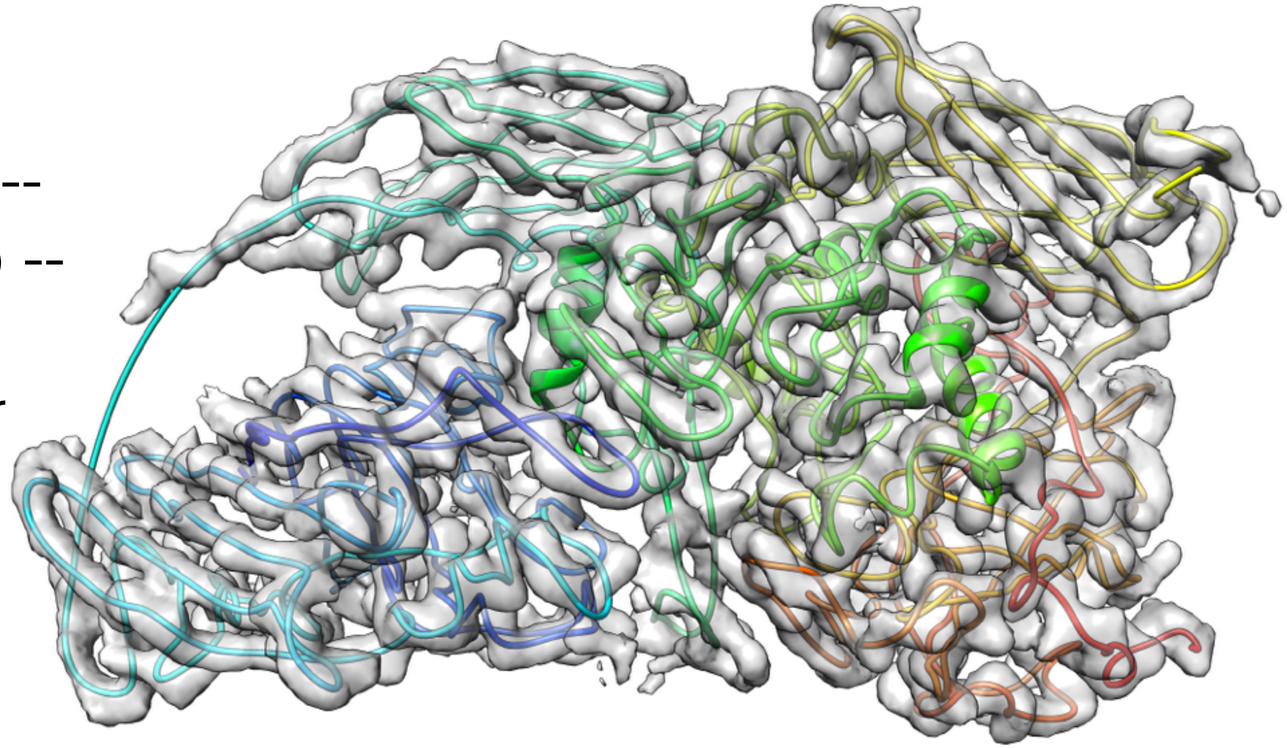
Pathwalker: Setting the termini

- In a terminal window run
  - e2pathwalker.py pseudoatoms.pdb --mapfile=map.mrc --output=path3.pdb --solver=lkh --overwrite --dmin=1 --dmax=10 --mapthresh=12 --mapweight=200 --edgefile=edge.txt **--start=91 --end=956**
  - Open path3.pdb file in chimera
  - Render as ball and stick



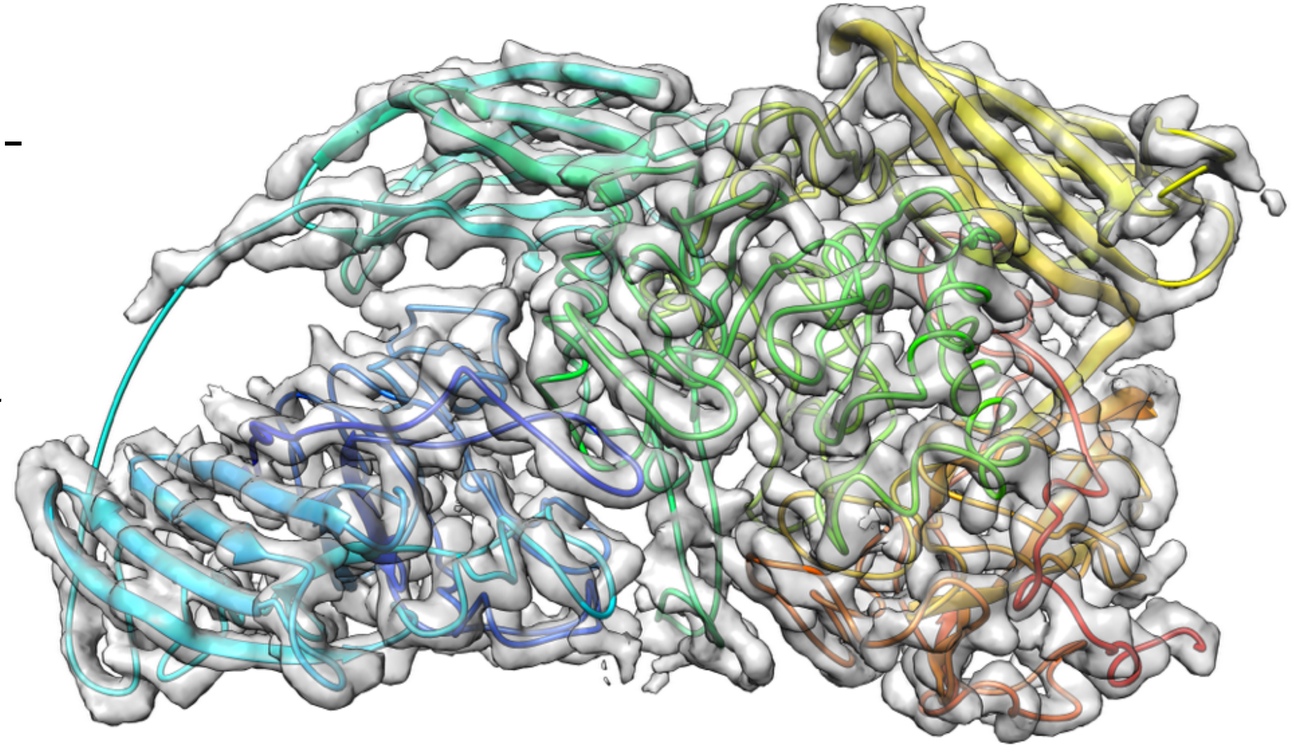
Pathwalker: Finding helices

- In a terminal window run
  - /Applications/EMAN2/examples/ e2pwhelixfit.py --mapin map.mrc --pdbin path3.pdb --output hlx.pdb --denthr 13 --mapwohelix map_nohlx.mrc --minlen 4 --lenthr 10
  - Open hlx.pdb file in chimera



Pathwalker: Finding sheets

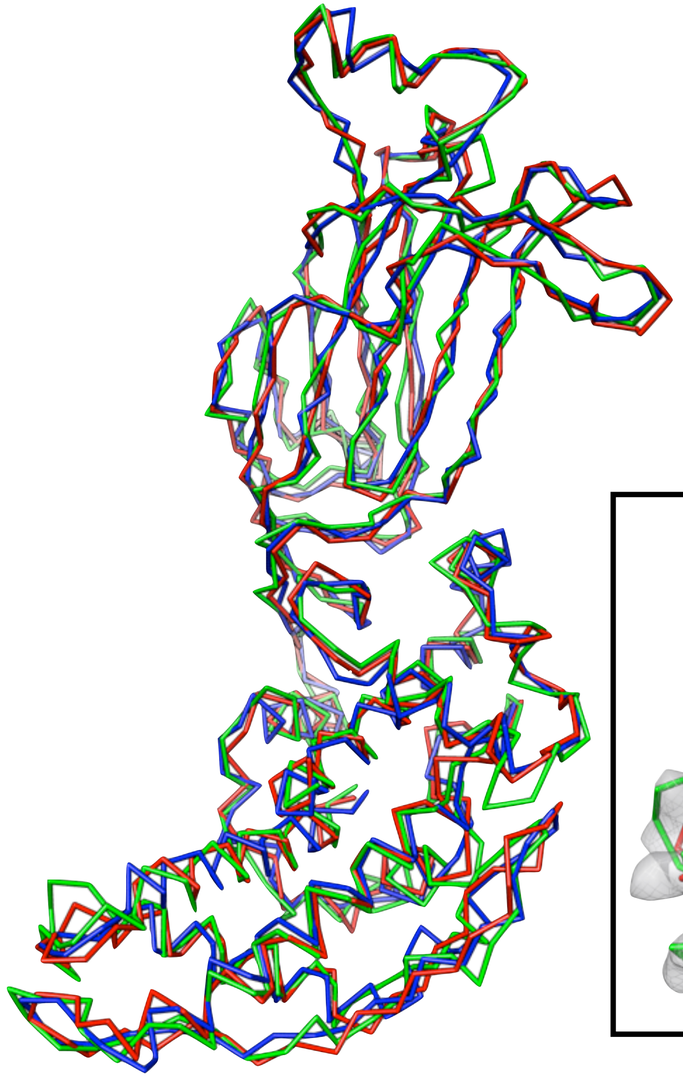
- In a terminal window run
  - /Applications/EMAN2/examples/ e2pwsheetfit.py --pdbin hlx.pdb --output sheet_0.pdb --nsht 30 --minlen 3
  - Open sheet_0.pdb file in chimera

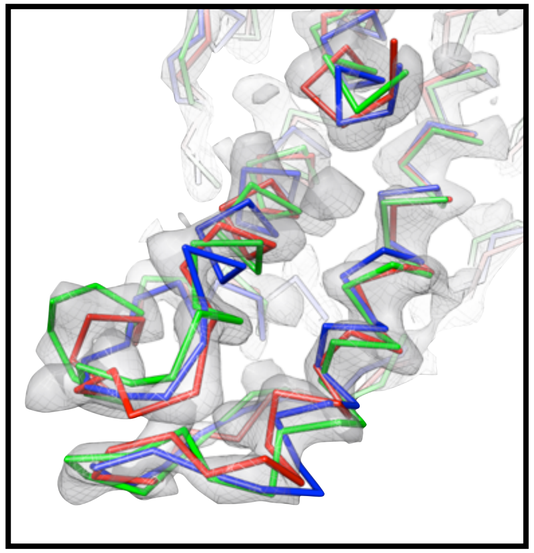


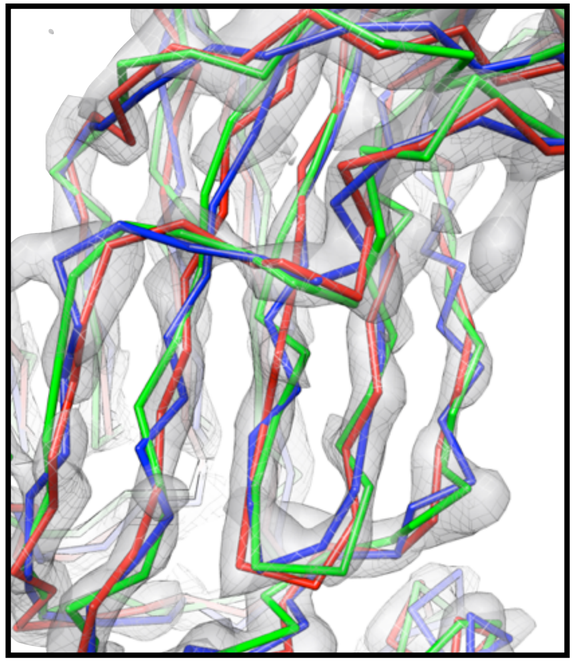
De Novo Modeling Utilities

	de novo	**Gorgon**	**Pathwalker**
Map requirements	full map	full map (<256^3)	segmented map
SSE	required	required	optional
SSE correspondence	required	required	none
Completion time	long (weeks)	short (0.5-1 day)	short (0.5-1 day)
Model accuracy	+++	++	+
User interface	varied software	graphical	EMAN2+Gorgon
Resolution range	3-5Å	3-7Å	3-6.5Å
Multiple models	no	partial	yes
Ease of use	difficult	easy to moderate	easy

Gorgon vs. Pathwalking







red= X-ray structure
blue= Gorgon mode
green= Pathwalker model

What Tools Should I Use?

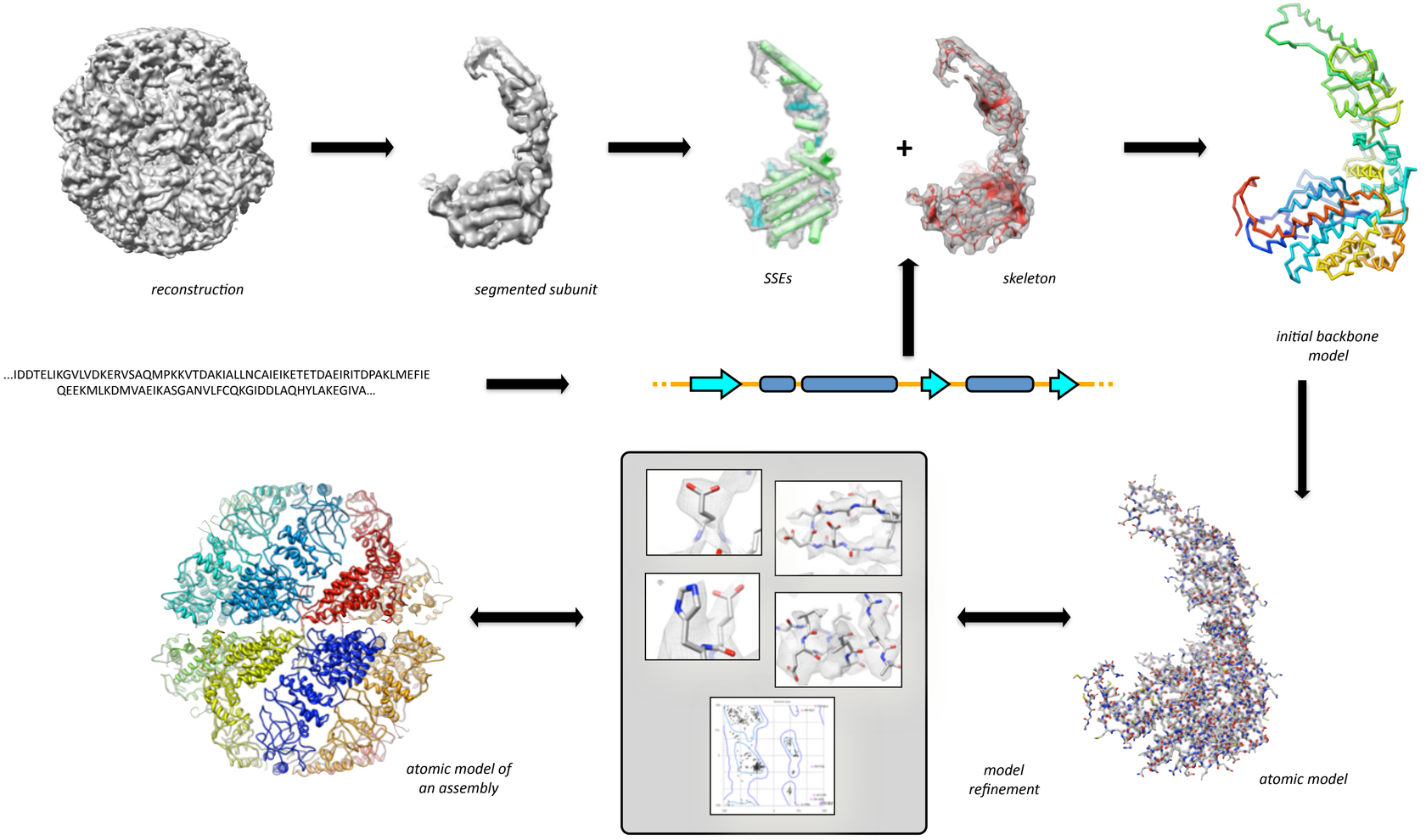
Gorgon

- Single protein subunit
- Between 3 and ~60 helices of varying sizes
- Well resolved density
- No size limit
- Interactive

Pathwalker

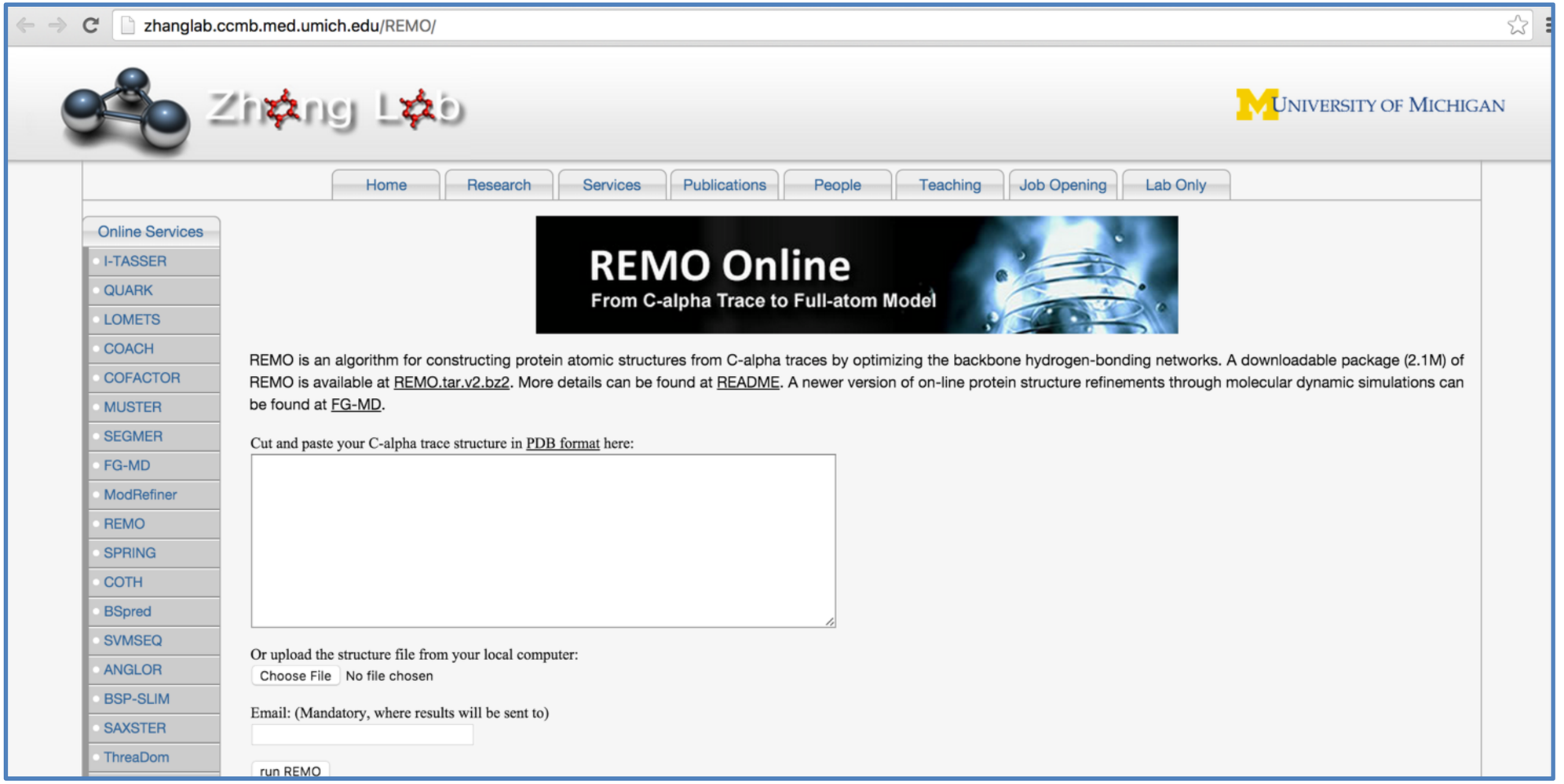
- Single or multiple subunits
- No secondary structure elements detected
- Between 3-7Å resolution
- Up to ~2000 amino acids
- Automated

All Atom Modeling



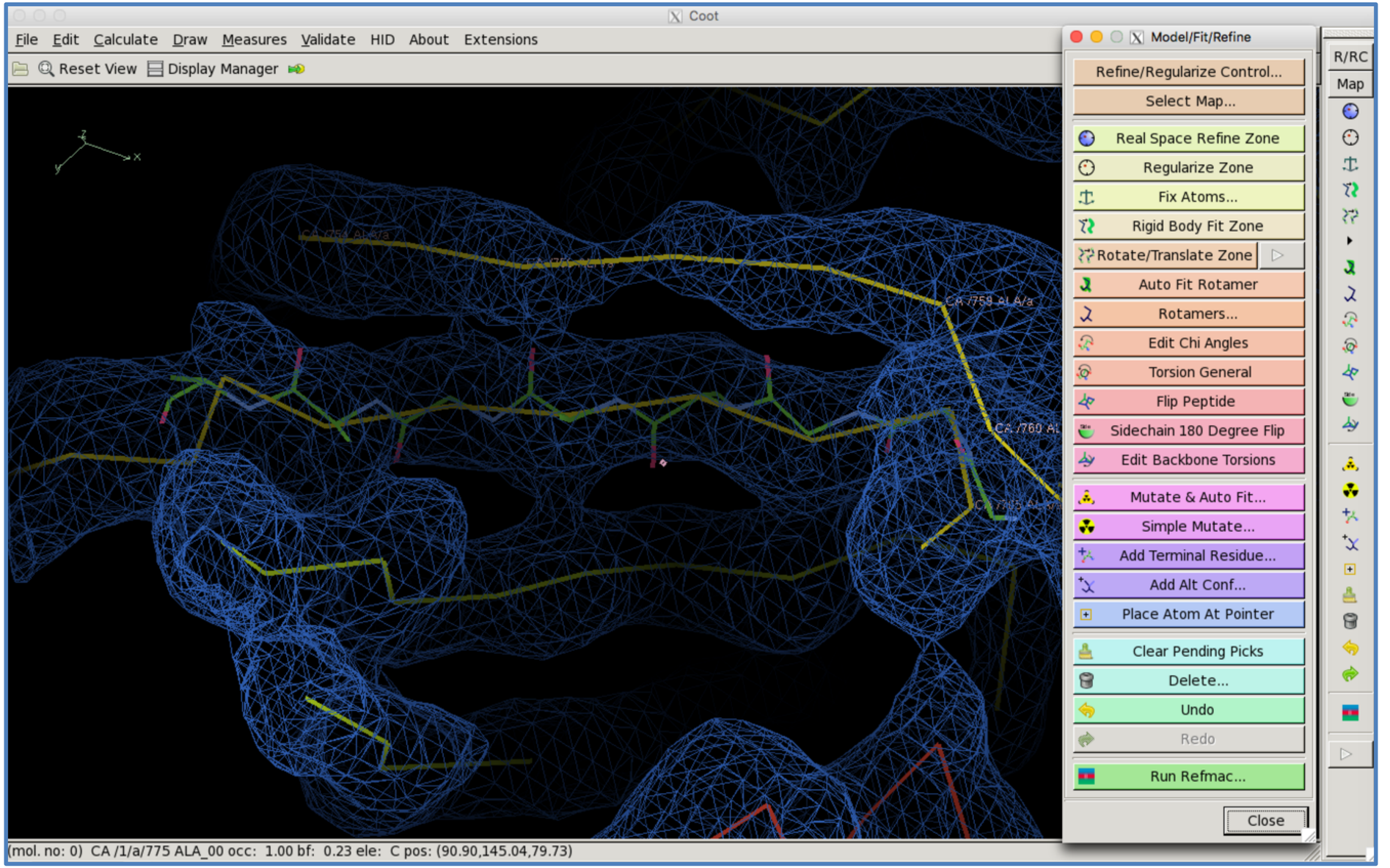
Baker ML, Zhang J, Ludtke SJ, Chiu W. Cryo-EM of macromolecular assemblies at near-atomic resolution. Nat Protoc. 2010 Sep;5(10):1697-708.

From C-alpha to All Atom



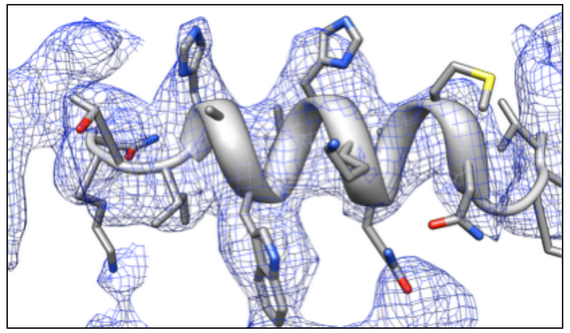
or phenix.pulchra sheet_0.pdb

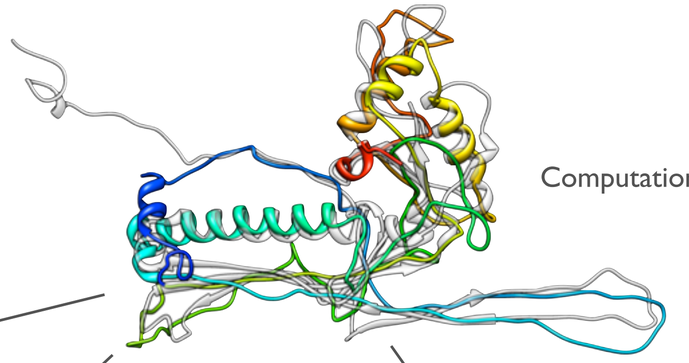
Coot


Coot
File Edit Calculate Draw Measures Validate HID About Extensions
Reset View
Display Manager
Model/Fit/Refine
Refine/Regularize Control...
Select Map...
Real Space Refine Zone
Regularize Zone
Fix Atoms...
Rigid Body Fit Zone
Rotate/Translate Zone
Auto Fit Rotamer
Rotamers...
Edit Chi Angles
Torsion General
Flip Peptide
Sidechain 180 Degree Flip
Edit Backbone Torsions
Mutate & Auto Fit...
Simple Mutate...
Add Terminal Residue...
Add Alt Conf...
Place Atom At Pointer
Clear Pending Picks
Delete...
Undo
Redo
Run Refmac...
Close
R/RC
Map
(mol. no: 0) CA /1/a/775 ALA_00 occ: 1.00 bf: 0.23 ele: C pos: (90.90,145.04,79.73)

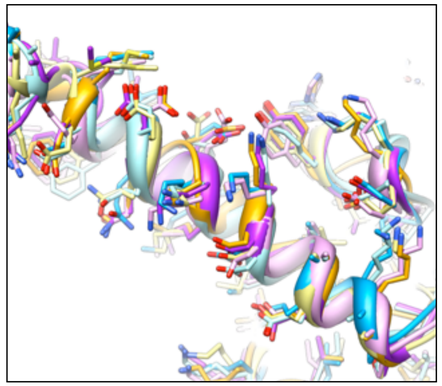
Model Refinement with Rosetta and Phenix

Bond distances /
Fit to density

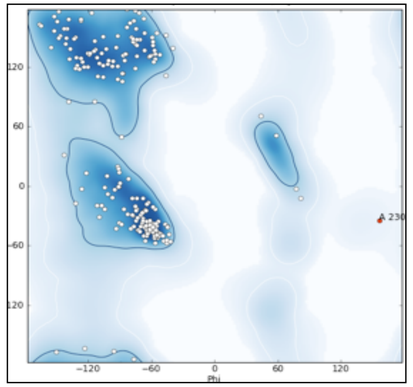




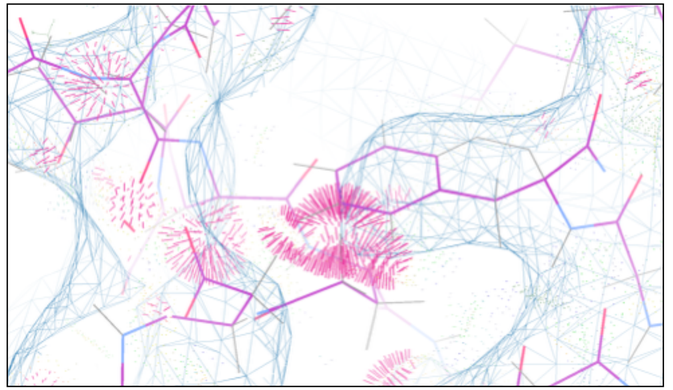
Computationally refined



Rotamer gallery



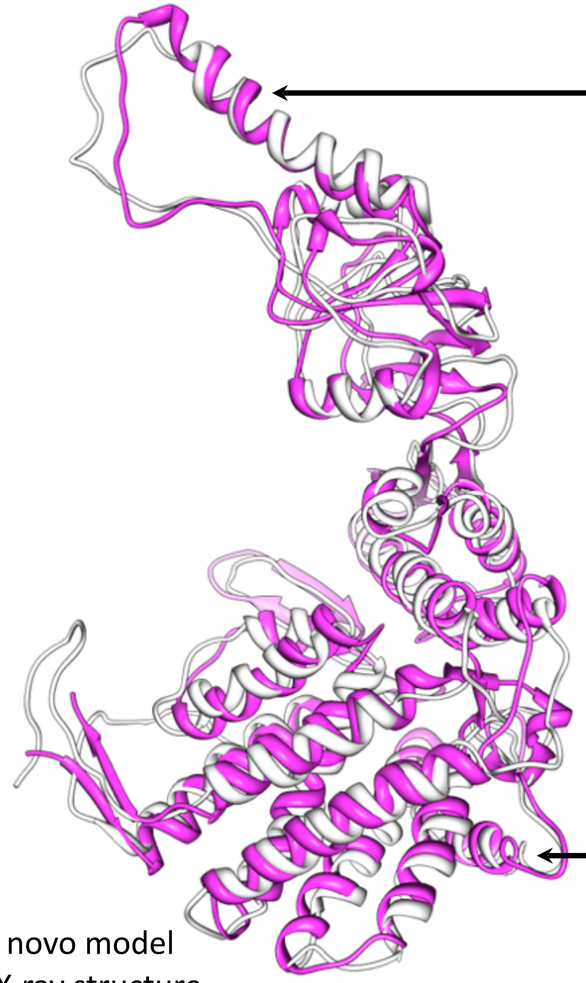
Ramachandran plot



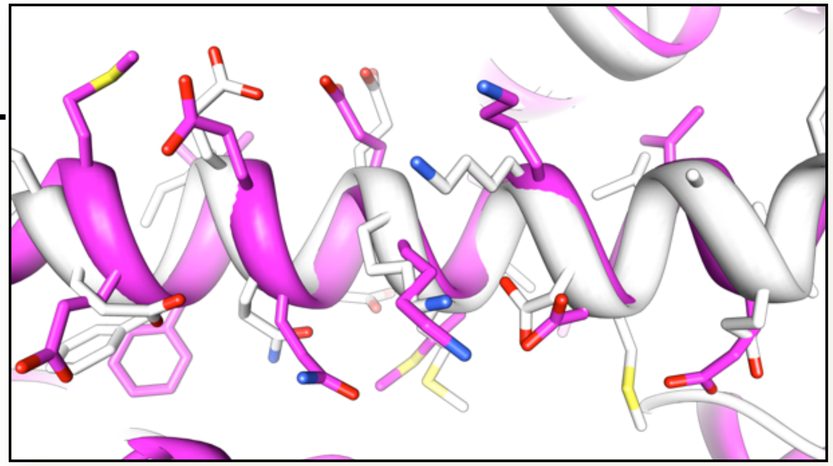
Clashes

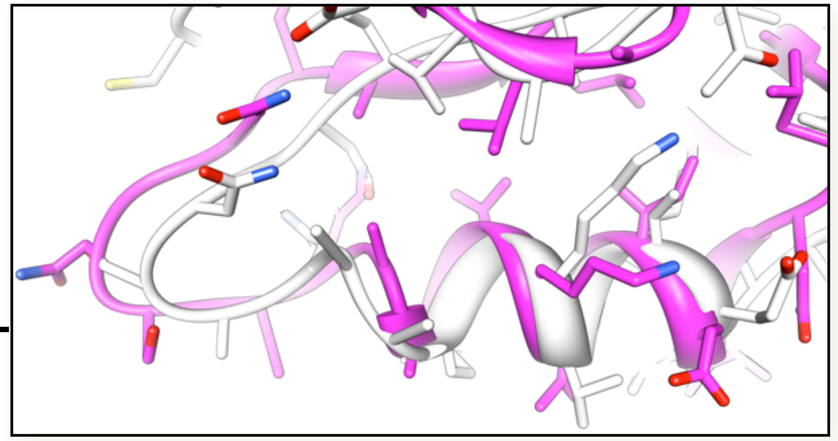
Baker, M. L., Hryc, C. F., Zhang, Q., Wu, W., Jakana, J., Haase-Pettingell, C., Afonine, P. V., Adams, P. D., King, J. A., Jiang, W., and Chiu, W. (2013). Validated near-atomic resolution structure of bacteriophage epsilon15 derived from cryo-EM and modeling. Proc Natl Acad Sci U S A 110, 12301-6.

Comparison to the X-ray structure



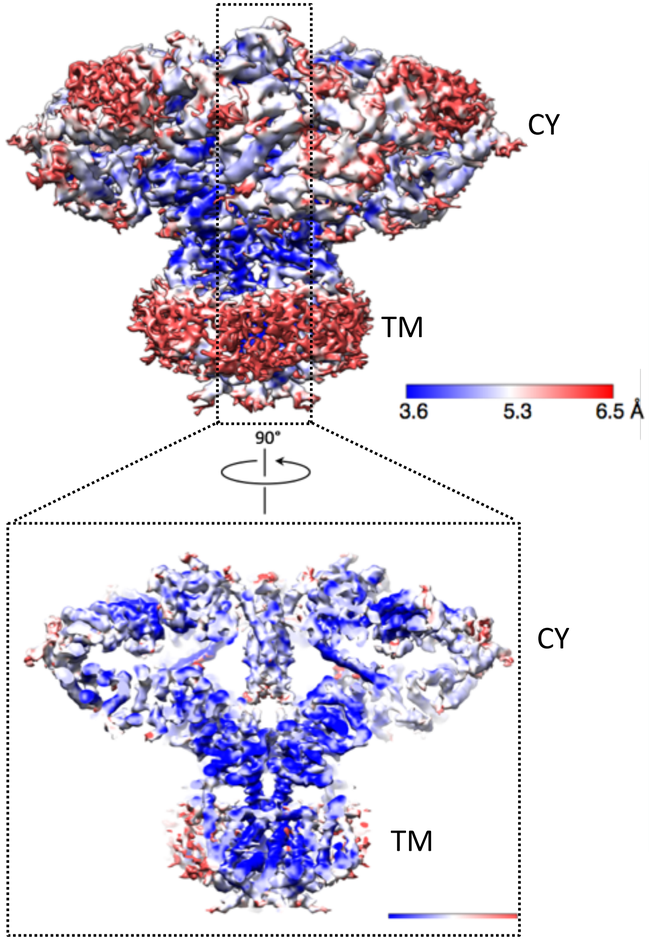
grey= de novo model
purple= X-ray structure

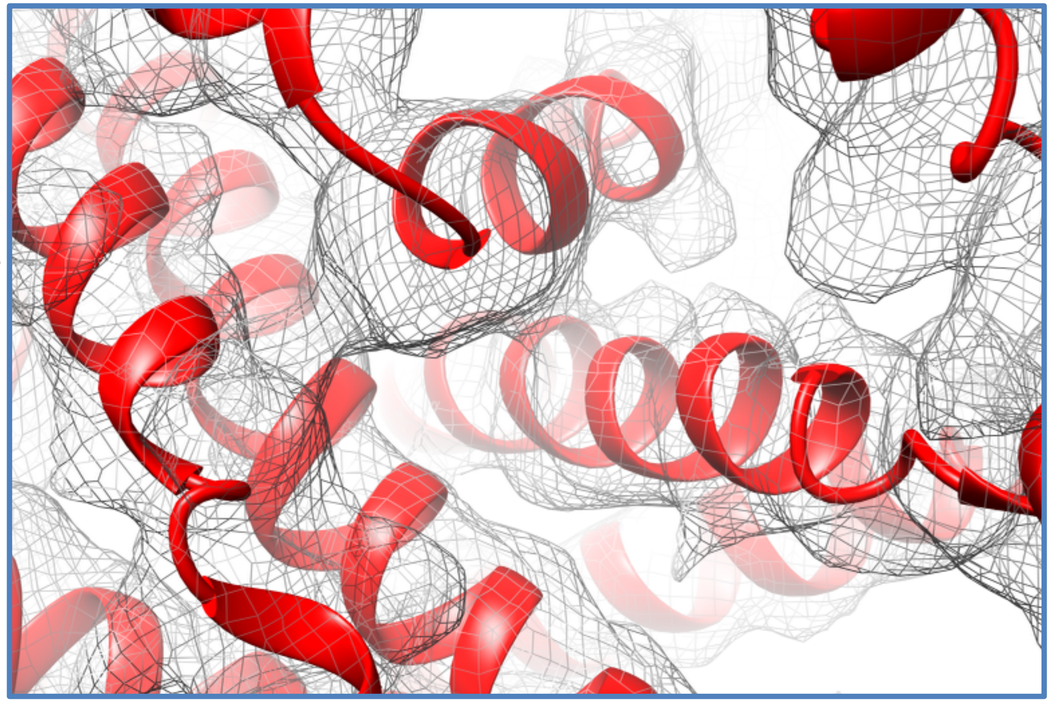




Zhang J, Baker ML, Schröder GF, Douglas NR, Reissmann S, Jakana J, Dougherty M, Fu CJ, Levitt M, Ludtke SJ, Frydman J, Chiu W. Mechanism of folding chamber closure in a group II chaperonin. Nature. 2010 Jan 21;463(7279):379-83.

Side Chain and Model Refinement





- Higher resolution features visible in TM, ILD, LNK, CTD and ARM3 domains
- Extend C-alpha domain models to full atom models using REMO
- Iterative real-space refinement in Phenix and manual optimization in COOT
    - optimize ramachandran plot, rotamer selection, clash minimization
- Model characterization and validation with Phenix and MolProbity

Acknowledgments

IP3R

- Mariah Baker (UTH)
- Guizhen Feng (UTH)
- Irina Serysheva (UTH)
- Zhou Wang (BCM)
- Steve Ludtke (BCM)
- Wah Chiu (BCM)

Pathwalker

- Mariah Baker (BCM,UTH)
- Ian Rees (BCM)
- Steve Ludtke (BCM)
- Muyuan Chen (BCM)
- Wah Chiu (BCM)

Gorgon

- Tao Ju (WUSTL)
- Sasakthi Abeysinghe, (WUSTL)
- Stephen Schuh (WUSTL)
- Derrick Burrows (WUSTL)
- Hang Dao (WUSTL)
- Mike Marsh (BCM)
- Ross Coleman (BCM)
- Tunay Durmaz (BCM)